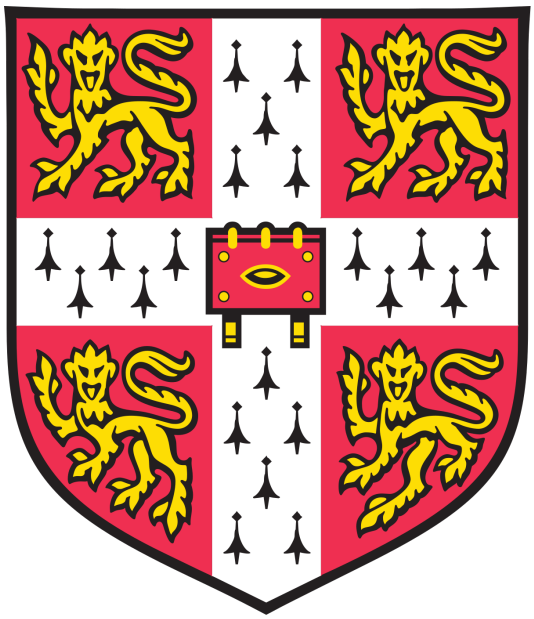
Chemical Mutagenesis for the Modulation and Study of Protein Function and Dysfunction within Alzheimer's Disease



Philip Richard Lindstedt

St. John's College

Department of Chemistry

University of Cambridge

This thesis was submitted for the degree of *Doctor of Philosophy*

November 2020

Declaration

This dissertation is a summary of research carried out under the supervision of Prof. Sir Christopher Dobson, Prof. Michele Vendruscolo and Dr. Gonçalo Bernardes in the Department of Chemistry at the University of Cambridge, between October 2017 and October 2020. The work described in the dissertation is my own and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements. It has not, either in part, or as whole, been submitted for a degree, diploma, or other qualification at any other university. The length of this dissertation (approx. 38,000 words) does not exceed the word limit set by the Degree Committee (60,000 words).

Philip Richard Lindstedt

November 2020

Chemical Mutagenesis for the Modulation and Study of Protein Function and Dysfunction within Alzheimer's Disease

Philip Richard Lindstedt

Abstract

Chemical mutagenesis is defined as the post-expression interconversion of amino-acid (AA) sidechains. This approach enabled the first ever point-mutations in proteins and has seen a resurgence of interest over the last two decades. This thesis reports on two disparate applications of chemical mutagenesis within the context of Alzheimer's disease (AD). First, exploring the capabilities of utilizing chemical mutagenesis to enable a structure-activity relationship study on a single-domain antibody (SdAb) beyond the 20 canonical AAs. The starting SdAb was designed to target the Amyloid-b peptide (Ab) and has a nascent ability to inhibit its aggregation, a key pathological process in AD. Utilizing the synthetically versatile non-canonical AA dehydroalanine (Dha), we explore diverse sidechains at key residues along the binding loop of the SdAb and identify a non-canonical sidechain that potently enhances the inhibition of Ab's primary nucleation. The second application focuses on the other major protein implicated in the pathology of AD: Tau. In AD, tau aggregates into ordered amyloid fibrils and contains a plethora of post-translational modifications (PTMs), many of which are speculated to have an impact on the pathogenesis of AD as well as other so-called tauopathies. In order to facilitate the accurate study of tau PTMs, an easily accessible method for their site- selective installation is needed. To this end, we again explore the utilisation of Dha to install accurate PTM mimetics for phosphorylation, lysine acetylation, and lysine dimethylation at diverse sites along tau. This approach produced homogenous products that accurately recapitulated known behaviours and allowed for the initial characterisation of poorly investigated PTMs. The characterisation of S199 phosphorylation, a known AD associated modification, through this approach provides the first evidence for its potential role in regulating tau's primary function of microtubule polymerisation and stabilisation.

Acknowledgments

First and foremost, I want to acknowledge the people that brought me to and supervised me at Cambridge. To Professor Sir Christopher M. Dobson, I am forever indebted for reading my initial email proposing me to join the Centre for Misfolding Diseases. To Professor Michele Vendruscolo, I am so grateful for your unwavering support and guidance in all of my scientific endeavours. And to Dr. Gonçalo Bernardes, I am forever indebted for your endless guidance on all things to do with protein modification, as well as your love of Seattle.

My mother, Julia Lindstedt, deserves more praise than I could ever hope to give. For supporting me all while my father succumbed to Alzheimer's, I owe her more than the world. You have been an inspiration to so many more than you know, and I always hope to live up to your expectations.

My partner in life, Emily Matthews, I am infinitely grateful for your love and support through the highs and lows of my research. I know I have leaned on you more than anyone else during difficult times over the last several years, and I would not be writing these words without you.

My sister Paulina Lindstedt, I am forever lucky to have you as friend throughout all of life's ups and downs. Thank you for always keeping me grounded the only way a sister can.

Dr. Francesco Aprile, thank you for your exceptional scientific guidance. From helping me purify Hsp70 in my master's degree to planning my future after my PhD, you have been there every step. I am forever indebted to you.

Dr. Pietro Sormanni, you have been a friend and inspiration for so many years. Your knowledge of both theory and experimentation is truly unrivalled. From holding my hand in the beginning of my experience with antibody engineering, to advising me today in the next steps of my career, I am indebted to you. Your dedication to your work is an inspiration for all.

Dr. Janet Kumita, you have been the most incredible supporter of me and everything I've done at my time in Cambridge. You've been a wonderful friend to me and always been there when I needed your help and for that I am forever grateful.

Dr. Faidon Brotzakis, thank you very much for sharing your molecular dynamics results with me and for our discussions on all things tau.

Dr. Sam Ness, thank you so much for your comradery. I was so unbelievably lucky to find such a genuine friend in my place of work, yet alone someone just as obsessed with all things to do surrounding antibodies. Your optimism and banter kept me going through many challenging times.

Ewa Klimont and Swapan Preet, thank you for providing me and the rest of the lab a never-ending supply of Aβ. We all appreciate your support in everything related to protein purification.

Robertinah Rakoto, you were an immense help during my work with the antibodies. Your thirst for knowledge was insatiable and contagious and drove me to be a better scientist.

Ross Taylor, your enthusiasm for chemistry drove me to explore my own horizons. I hope you are lucky enough to find similar people to challenge your perspective's during your doctoral research.

I was very fortunate to make a truly lifelong set of friends during my first year at St. John's College. To Patrick Chitwood, Lisa Hawkins, Fynn McKay, Gaelen Stanford-Moore, and Tom Mosley, I am so privileged to have a group of immensely talented and genuine friends. I have drawn inspiration from you all and am so excited to see what lies ahead for all of us.

This thesis is dedicated to the two people who I wish could have read it the most. First and foremost, to my father, Per Lindstedt, who Alzheimer's took from us far too early. I miss you so terribly every day. And second to Chris Dobson, who supported every one of my ideas and consoled me with stories of his father. You were an inspiration to so many more than I could list, as a scientist and human being, and I will miss you more than I ever had the chance to say.

List of Abbreviations

AA	Amino acid
aaRS	Aminoacyl-tRNA-synthetase
AD	Alzheimer's disease
ADC	Antibody drug conjugate
AFM	Atomic force microscopy
ANS	8-Anilinonaphthalene-1-sulphonic acid
APP	Amyloid precursor protein
Aβ	Amyloid-β
BLI	Biolayer interferometry
BME	β-mercaptoethanol
BSA	Bovine serum albumin
CD	Circular dichroism
CDR	Complementarity determining region
CMD	Centre for Misfolding Diseases
CNS	Central nervous system
DBHDA	2,5-dibromohexanediamide
DesAb	Designed antibody
Dha	Dehydroalanine
DMF	*N,N*-dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
Fab	Antigen-binding fragment
Fc	Crystallisable fragment
FcRn	Neonatal Fc receptor
FDC	Follicular dendritic cell
Fv	Variable fragment
GSK3β	Glycogen synthase kinase-3β

HcAbs	Heavy-chain antibodies
IDP	Intrinsically disordered protein
IEC	Ion-exchange chromatography
IEDDA	Inverse-electron demand Diels-Alder
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k_+	Elongation rate constant
k_-	Fragmentation rate constant
k_2	Secondary nucleation rate constant
K_D	Equilibrium dissociation constant
k_n	Primary nucleation rate constant
k_{off}	Off-rate constant
k_{on}	On-rate constant
LB	Luria broth
LC-MS	Liquid chromatography-mass spectroscopy
mAb	Monoclonal antibody
MAPK	Mitogen-activated protein kinase
MBD	Microtubule binding domain
MDBP	methyl 2,5-dibromopentanoate
MESNA	2-mercaptoethanesulfonate
MRE	Mean residue ellipticity
mRNA	Messenger RNA
MT	Microtubule
MWCO	Molecular weight cut-off
NAC	Non-amyloid-β component
NAL	*N*-acetylneuraminic lyase
NCL	Native chemical ligation
NEB	New England Biolabs
NFTs	Neurofibrillary tangles
PBS	Phosphate-buffered saline
PCR	Polymerase chain-reaction
PD	Parkinson’s disease
PDB	Protein data bank

PEG	Polyethylene glycol
PERx	Proximity enabled reactive therapeutics
PET	Positron emission tomography
PHF	Paired-helical filaments
PKU	Phenylketonuria
PMSF	Phenylmethylsulfonyl fluoride
PRD	Proline rich domain
PTM	Post-translational modification
RMSF	Root mean square fluctuations
RNA	Ribonucleic acid
RT	Room temperature
SAR	Structure activity relationship
scFv	Single-chain variable fragment
SDS PAGE	Sodium dodecyl-sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SHM	Somatic-hypermutation
SM	Small molecule
TB	Terrific broth
TEM	Transmission electron microscopy
ThT	Thioflavin-T
T_m	Melting temperature
tRNA	Transfer RNA
UAA	Unnatural amino acid
UV	Ultra-violet
V_H	Variable heavy domain
VHH	Variable heavy-chain fragment
V_L	Variable light domain

Table of Contents

Chapter 1: Introduction

1.1 Protein aggregation and disease

1.1.1. Protein folding

At the most fundamental level, the main requirement for the propagation of living systems is its components' ability to reliably self-assemble. From this deceivingly simple capability, life is able to build the complex systems necessary for everything from metabolism, to replication, and all of the other intricate functions leading to the great diversity that life has achieved. Essential to the self-assembly required for life is the ability of proteins to fold into their native three-dimensional structure[1]. It is hard to overstate how profound it is that proteins can, by and large, fold on their own incredibly rapidly (on the order of μs-ms)[2]. While native states almost always correspond to the most stable structure of a protein under physiological conditions[3], within the cell proteins are generated as a linear polypeptide from the exit tunnel of the ribosome. In the case of the small 110 residue protein Barnase, from estimates of main-chain entropy per residue, the number of conformations per residue in this initial random-coil state was calculated to be 2.57, when extended to the whole protein that becomes 2.57^{109} ($\sim 10^{44}$) number of states[4]. A random search through this conformational space to the native state would take much longer than the lifetime of the universe. This observation is famously known as Leventhal's paradox[4]. From observations such as this, it became clear that the folding of proteins, at least for those capable of spontaneously folding, is largely encoded by its amino acid sequence. Through the implementation of a myriad of experimental and theoretical techniques, it became clear that within the sequence of proteins was encoded a free-energy landscape to guide the folding of the protein[5] (**Figure 1.1a**). The landscape is shaped by the fact that, on average, native-like interactions between residues are more stable than aberrant interactions, and once a critical few interactions have been made to create a folding nucleus, then the remaining structure rapidly condenses into the native state[6]. This is at least the case when folding is performed correctly.

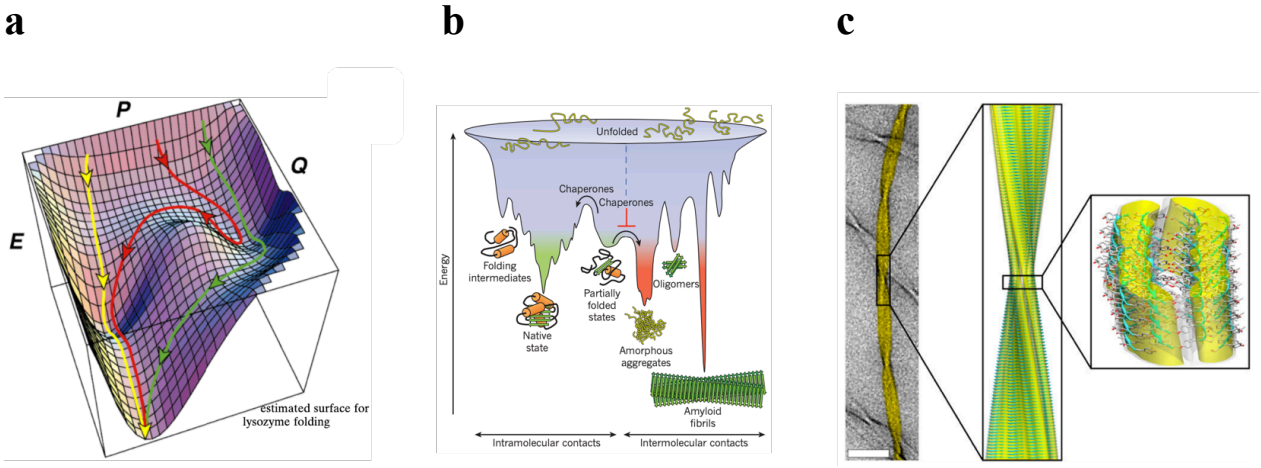


Figure 1.1. Protein folding and the amyloid state. (a) The averaged effective energy surface of the folding of lysozyme based on experimental data with three folding trajectories: fast folding (yellow), slow folding crossing a high energy barrier (green), and a slow folding state that returns to a less folded state before following the fast folding trajectory (red). *E* represents the free energy, *P* is a measure of the conformational space, and *Q* is the number of native contacts. (b) Illustration depicting the stability of amyloid fibrils and their intermediate aggregate's free energy compared to the native state and a possible point of intervention molecular chaperones can take to inhibit the process. (c) TEM image of the amyloid fibrils from an 11-residue peptide (*left*) and close-up view of the magic-angle spin NMR atomic resolution structure of the fibrils (*centre/right*). Figures modified with permission from references [2], [8], and [10], respectively.

1.1.2. The amyloid state

While the amino acid sequence of proteins can guide the folding process, there are many perturbations *in vivo* that can either interfere with the initial folding or cause partial or complete unfolding of a protein[7]. Functionally compromised proteins are closely monitored by a proteostasis system that is highly conserved through all domains of life[8]. Within this system, there are a plethora of entities that either assist with the initial folding, refolding, or degradation of irrevocably lost proteins. One increasingly recognised function of this system is the inhibition of the amyloid state, an intractable and highly stable protein structure[9]. (**Figure 1.1b**).

Interestingly, regardless of the protein or peptide of origin, amyloid structures share many similar traits. This state is characterised by the formation of long (often micron-length) fibrils possessing a cross-β structure where the β-strands are oriented perpendicular to the fibril axis[10] (**Figure 1.1c**). Remarkably, once adopted by a protein, the amyloid state can be even more thermodynamically favourable than the native state (**Figure 1.1b**). In contrast to how the

native state of a protein is largely coded in its amino-acid sequence, there are many observations that suggest that the amyloid state is an innate structure all polypeptides can adopt. This hypothesis was inspired by observations that conditions could be found for almost any normally globular protein to convert to amyloid fibrils. An initial striking demonstration of this was with muscle myoglobin by Fandrich and co-workers[11]. In its native state, muscle myoglobin is a highly soluble and α-helical rich protein, and all of its partially folded states are also predominantly α-helical[12]. However, the researchers found that if the apo form of the protein (lacking haem) was incubated in 50 mM sodium borate buffer, pH 9.0 and at 65 °C, it resulted in the deposition of long ordered fibrillar structures with the characteristic signatures of a cross-β structure by X-ray diffraction and FTIR. The identification of conditions for the conversion of numerous other normally soluble proteins into amyloid fibrils supports this observation[13-15]. However, perhaps most convincing of the theory of universal amyloid formation is the ability of single amino acids to form the structure. In a seminal study by Adler-Abramovich and co-workers, the pathological deposition of amino acid *L*-phenylalanine was studied in detail in the context of the disease phenylketonuria (PKU)[16]. PKU is an autosomal recessive disorder causing an inability to metabolise *L*-phenylalanine, leading to the accumulation of *L*-phenylalanine to millimolar levels throughout several body tissues, including the brain[17]. The researchers found that under physiological conditions *in vitro* this high concentration leads to the deposition of highly ordered supramolecular assemblies with all of the hallmarks of amyloid fibrils mentioned previously. How is it possible that so many disparate sequences can adopt this structure? While in the native state side-chain interactions are the key players in defining the main-chain fold, it has become clear that the amyloid structure is driven by main-chain interactions that are common to all polypeptides. Indeed, it has been shown that it is the longitudinal backbone hydrogen bonding network along the fibril axis that is the main contributor to the stability of these structures[18]. However, the kinetic barriers to accessing the amyloid state are heavily influenced by the amino acid sequence and conditions.

These examples of aberrant amyloid formation by natural protein sequences and peptides led many to believe that nature had evolved to select for sequences, conditions and systems to suppress the formation of this universal structure. However, nature has a remarkable ability to utilise even extreme physiochemical phenomena. It is now clear that functional amyloids are pervasive through all domains of life[19]. Within mammals, this is exemplified by the protein Pmel17, which is responsible for the polymerisation of melanin[20]. Fowler and co-

workers demonstrated through histochemistry that mammalian melanosomes contain a significant level of amyloid, suspected to arise from Pmel17 through genetic studies[21,22]. Indeed, when Pmel17 was recombinantly expressed and isolated it was then found to form amyloid fibrils at a remarkable rate (rate-limited by the experimental apparatus). The function of the amyloid nature of Pmel17 is believed to be as a scaffold to template the polymerisation of melanin and sequester the reactive and toxic melanin precursors from the cellular milieu. While the utilisation of the amyloid fold for cellular activity was a paradigm shift in the protein misfolding field, it is now established that that these functional amyloids have a multitude of features that have evolved to avoid the more malicious traits of these structures.

1.1.3. Pathological protein misfolding

As alluded to in the above section, amyloid formation has most famously been associated with disease processes. Protein misfolding can bring about disease through two broad mechanisms: loss of function or gain of toxicity. In the loss of function mechanism, impaired folding results in either enhanced degradation or improper trafficking of the protein, leading to a reduction of the level of functional protein to perform its normal role. This is the case in diseases like cystic fibrosis and early onset emphysema[23,24]. The majority of protein misfolding diseases, however, fall into the gain of toxicity category. Over the last several decades thirty-seven proteins and peptides have been found to form amyloid structures in a variety of human maladies (see **Table 1.1** for examples)[25]. The symptomatic manifestations of these diseases vary widely, from dementia, to diabetes, to cardiomyopathy, and depend entirely on both the identity of the misfolding protein/peptide and the location of the amyloid deposits.

Aggregating peptide/protein	Number of residues	Associated diseases	Native structure
Neurodegenerative diseases			
Amyloid-β	40 or 42	Alzheimer's disease	Intrinsically disordered
Tau	352 - 441	Alzheimer's disease, Pick's disease, frontotemporal dementia, chronic traumatic encephalopathy, ganglioglioma, progressive supranuclear palsy, ~8 more diseases	Intrinsically disordered
α-synuclein	140	Parkinson's disease, Parkinson's disease with dementia, dementia with Lewy bodies, multiple system atrophy	Intrinsically disordered
Transthyretin	127	Senile systemic amyloidosis, familial amyloidotic polyneuropathy, familial amyloid cardiomyopathy (multiple disease categories)	All-β, prealbumin like
Huntington exon 1 (HttEx1)	103-187	Huntington disease	Intrinsically disordered
Non-neuropathic systemic amyloidosis			
Immunoglobulin (Ig) light chains or fragments thereof	~100	Light-chain amyloidosis	All-β, Ig like
Lysozyme	130	Lysozyme amyloidosis	α and β, lysozyme fold
β_2-microglobulin	99	Dialysis related amyloidosis	All-β, Ig like
Non-neuropathic localised amyloidosis			
Islet amyloid peptide	37	Type-II diabetes, insulinoma	Intrinsically disordered
Medin	50	Aortic medial amyloidosis	Intrinsically disordered
Galectin 7	136	Lichen amyloidosis, macular amyloidosis	All-β, concanavalin A-like lectins
Pulmonary surfactant-associated protein C (SP-C)	35	Pulmonary alveolar proteinosis	All-α, transmembrane helical fragment

Table 1.1. Examples of human diseases associated with amyloid aggregation.

Genetic studies of this disparate class of diseases reveal interesting insights into the perturbations that can lead to amyloidopathies (**Figure 1.2**). From gene duplication or

expansion, to RNA splicing alterations, to structural destabilisation, these types of studies have revealed that amyloidopathies can arise from mutations that cause malfunctions at every stage of the genesis and life of a protein. For large errors in the DNA replicative machinery, one of the most well studied cases is with the duplication and triplication of the α-synuclein gene (*SNCA*) in certain cases of familial Parkinson's disease. One of the hallmarks of Parkinson's disease is the deposition of amyloid plaques of the protein α-synuclein, leading to a progressive loss of dopaminergic neurons[26]. Triplication of the *SNCA* gene was observed in a Swedish-American family and resulted in an autosomal dominate early-onset form of Parkinson's with Lewy body dementia[27]. In contrast, duplication of the *SNCA* gene in a French family led to an autosomal dominant form of the disease but with a progression pattern resembling idiopathic forms and no cognitive impairments[28]. Analysis of the increased mRNA and soluble protein concentrations of brain tissues from these families revealed that an increase of *SNCA* gene copies dose dependently increases the expression of α-synuclein[27]. This correlation between expression of the amyloidogenic protein and disease severity is indicative of how errors in the larger replicative machinery at the DNA level can induce or exacerbate amyloidopathies. More subtle genetic alterations can also lead to the onset of amyloidopathies. This is the case for one of the most thoroughly studied misfolding diseases: Transthyretin amyloid disease. Normally, transthyretin is found in human serum as a stable homo-tetramer functioning as a transporter for thyroxine and retinol. More than 80-point mutations have been identified within the transthyretin gene that lead to the pathogenetic deposition of the amyloid form of the protein, invariably ending in fatal complications[29]. Through many biophysical studies, the root-cause of amyloid formation from these mutations was determined to be the destabilisation of the tetramer structure[30-33]. The dissociated monomer has a greatly diminished kinetic barrier to adopting the amyloid state compared to higher ordered forms, and the shift in binding equilibrium brought about by these point-mutations is enough to produce a critical concentration of unstable monomer to initiate amyloid formation. Unique to transthyretin amyloid disease, these thorough mechanistic studies did more than elucidate its aetiology, they laid the foundation to guide a therapeutic approach. Kelly and co-workers used the knowledge of the destabilisation of the tetramer to screen for small-molecule stabilisers of the pathogenic mutations[30]. This work has culminated with the first disease modifying therapy of a protein misfolding disease on the market that has gone on to save thousands of lives already[34].

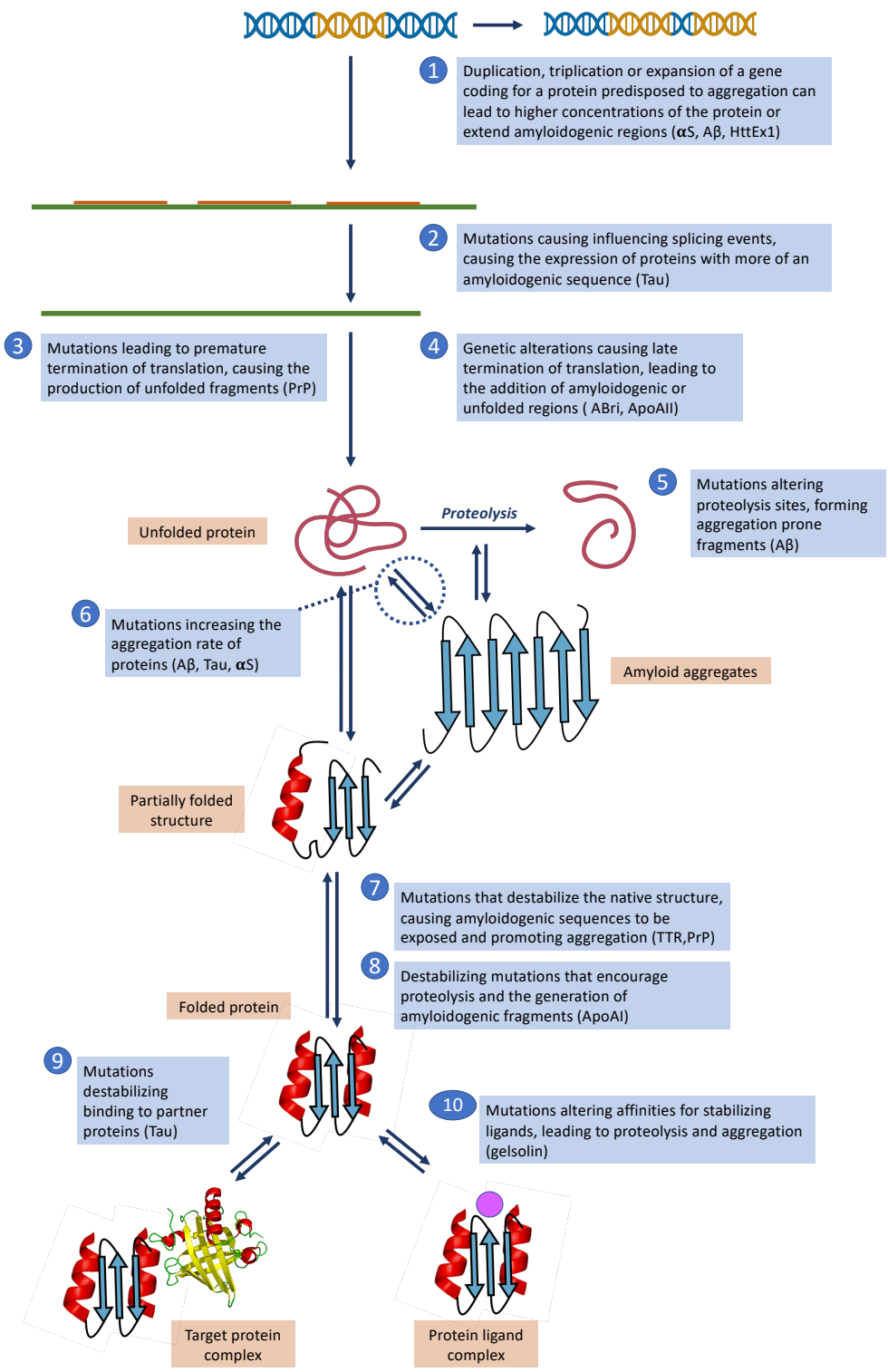


Figure 1.2. Mechanisms of action by which genetic alterations cause protein misfolding. Numbers 1-10 show different examples of aggregating proteins and the genetic alterations that encourage their misfolding.

While these genetic alterations explain how amyloid formation can be spurred, how do amyloid structures themselves cause disease? In the early days of the protein misfolding field, the pathological culprit was thought to be the amyloid fibrils themselves. In certain non-neurological amyloidopathies this is indeed the case, where kilograms of fibrils can be deposited, thereby leading to organ dysfunction[35]. However, for the majority of protein misfolding diseases, this turned out to be an overly simplistic and incomplete explanation for the pathogenicity of these structures. Over the decades it has become clear that most of these diseases have a very multifaceted mode of toxicity, interacting with a wide array of cellular and physiological systems. For the toxicity directly associated with misfolded proteins, it is increasingly accepted that the primary pathogenic species are the on-pathway oligomers that are generated throughout the aggregation process. This mechanism was initially hypothesised upon the observation that early stage non-fibrillar aggregates exhibited substantial toxicity to cultured cells, while fully fibrillar species were harmless (**Figure 1.3a**)[36]. Further insight was gleaned from this particular study because the proteins investigated were not disease associated; they were normally globular proteins that only form amyloid structures under the appropriate *in vitro* conditions, and yet, their on-pathway aggregates still were highly toxic, implying that oligomer mediated cell death is a potentially universal mechanism of toxicity across amyloidopathies. This hypothesis has been supported through a multitude of clinical observations[37-39]. For example, in transthyretin misfolding disease the onset of symptoms corelates with the deposition of the protein in a non-fibrillar form histologically distinct from later stage amyloid structures[40]. Follow up *in vitro* studies confirmed that low mass (<100 kDa) assemblies were the most cytotoxic species in the transthyretin aggregation process[41].

Oligomer mediated cell-death has been determined to largely arise through non-specific membrane disruption. The physiochemical characteristics of oligomers facilitating this disruption has been well characterised using the model protein HypF-N. HypF-N is an *E. coli* derived protein that is particularly well suited to this type of mechanistic study, this is because HypF-N can readily form an array of pre-fibrillar species that persist long enough to enable their study, unlike the typical highly transient species in other misfolding pathways[42]. In particular, conditions were found to produce two distinct types of stable oligomers with drastically different toxicities (**Figure 1.3b**)[43]. Initial characterisation of the two oligomer forms, toxic and benign, through typical tinctorial tests and atomic force microscopy (AFM) revealed very similar morphologies between the two. To perform a more in-depth investigation of the oligomer structures, a panel of 18 single cysteine mutants of the protein were expressed

and labelled with *N*-(1-pyrene)maleimide (PM). PM acts as a probe for proximity, because when two PM groups are within 10 Å of each other they form exited-state dimers, giving rise to a new fluorescence emission in the 430-470 nm region. This facilitated the investigation of the packing of the core of the oligomers, which revealed that in the toxic variant the core is much less tightly packed, and that the hydrophobic residues that are buried in the native structure and in the non-toxic oligomers are exposed on the surface. This was corroborated by measuring the binding of the different oligomers to the dye 8-anilinonaphthalene-sulfonate (ANS), which experiences a marked increase in fluorescence upon interacting with solvent-exposed hydrophobic clusters. This surface exposed hydrophobicity of oligomers appears to be the key characteristic that enables membrane disruption. Further structure-toxicity related studies of HypF-N oligomers have elucidated that, along with increased surface hydrophobicity, the toxicity of oligomers is inversely proportional to their size (**Figure 1.3c**)[44]. Supporting this general mechanism for misfolding diseases, nuclear magnetic resonance (NMR) studies of α-synuclein oligomers of various toxicities has elucidated the exact lipophilic structures that need to be present to mediate membrane disruption[45]. Taken altogether, these genetic and biophysical studies of protein misfolding over the last several decades have begun to create a wholistic picture of the pathogenic processes behind this diverse class of diseases.

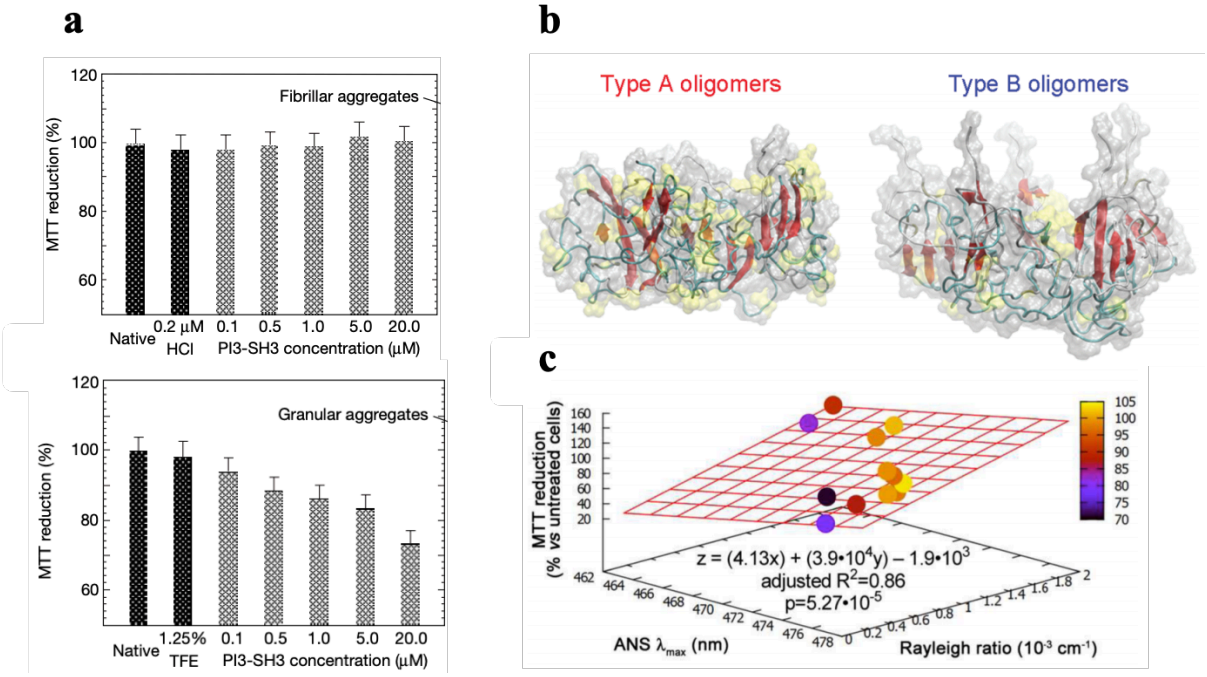


Figure 1.3. Toxicity of on-pathway oligomeric species. (a) Comparison of the toxicities of fibrillar aggregates (above) and granular (oligomeric, below) aggregates of the SH3 domain of

bovine PI3-kinase measured by MTT reduction. (b) Schematic representations of the structural differences of toxic (type A, red) HypF-N oligomers and non-toxic ones (type B, blue). Surface exposed hydrophobic side chains are highlighted in yellow. Structures were determined using a variety of biophysical techniques. (c) 3D plot displaying MTT reduction values (toxicity, z-axis), ANS λ_{max} (hydrophobicity, x-axis), and Rayleigh ratio (size, y-axis) of 12 different oligomeric variants of HypF-N. Figures modified with permission from references [36], [44], and [45], respectively.

1.1.4. Alzheimer's disease: The amyloid-ß peptide

Protein misfolding diseases are most prevalent in neurodegenerative disorders, in particular dementias, and none more famously than Alzheimer's disease (AD). Indeed, AD is poised to become a plague in the 21st century, with predictions that more than 140 million people will be afflicted with the disease by 2050[46]. AD, along with Parkinson's disease (PD), is typically a sporadic disease with no clear cause of onset. Although there are hereditary forms of the disease that shed some insight into its aetiology. However, the greatest risk factor by far for the disease is age, hence the predicted dramatic rise of cases in the 21st century due to the increasingly aged population of developed countries.

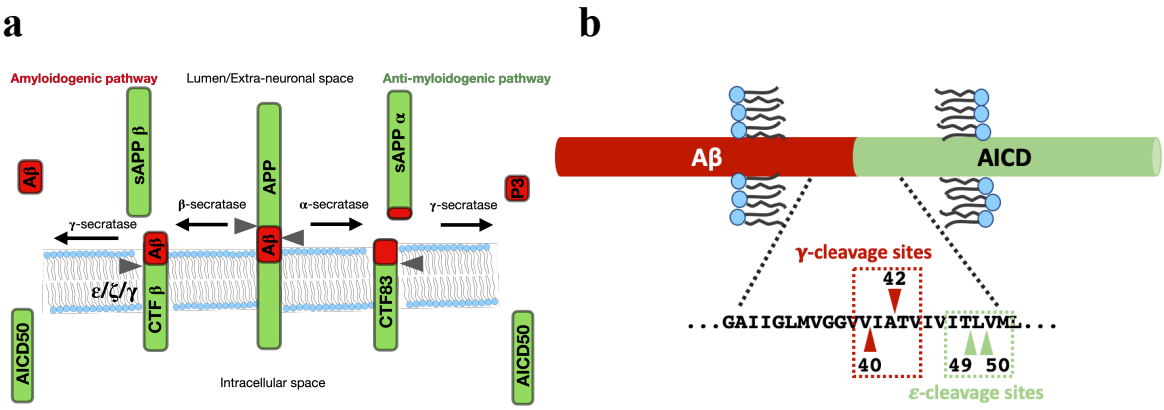


Figure 1.4. Aβ production. (a) Illustration of the amyloidogenic and anti-myloidogenic pathways of APP processing. (b) Cleavage sites of presenilin leading towards the $A\beta_{40}$ and $A\beta_{42}$ product lines and the preceding epsilon cleavage sites.

Post-mortem, the brain tissue of AD patients is characterised by the presence of two different proteinaceous inclusions: extracellular amyloid plaques and intracellular neurofibrillary tangles (NFTs)[47]. The morphological differences of these structures also reflect unique compositions, the former is largely composed of truncated peptides derived from the amyloid precursor protein (APP), predominately the $A\beta_{42}$ peptide, and the latter are formed by the microtubule associated protein tau. While both of these structures contribute to the disease progression of Alzheimer's, there is considerable evidence that the accumulation of $A\beta_{42}$, and

related peptides, is the beginning of the pathogenic cascade. A major observation supporting this is that all dominant mutations that lead to AD are either in the APP gene, or in the proteases that processes APP[48]. APP is a highly conserved type-1 transmembrane protein (luminal/extracellular amino terminus and cytosolic carboxy terminus) with a largely unknown primary function. While it is expressed in many tissues, it is found at the highest levels in neurons[49]. APP is processed by a succession of different proteases that can be classified into two main pathways: the amyloidogenic pathway and anti-myloidogenic pathway (**Figure 1.4a**). The two paths begin their divergence when the initial cleavage is carried out by α or ß-secretase. α-Secretase will cleave approximately in the middle of the Aβ region of APP, thereby avoiding the production of pathological species altogether, while the β-secretase (BACE1) cleaves deeper into the luminal domain, opening up the possibility of the production of highly amyloidogenic peptides[50]. The final release of the amyloidogenic peptides from the β-processed fragment (CTFβ) is then mediated by the γ-secretase, which is a complex consisting of 4 subunits, with the catalytic unit being the membrane-embedded aspartyl protease presenilin[51]. Presenilin processes CTFβ in a complex stepwise cleavage mechanism within the transmembrane domain (**Figure 1.4b**). Akin to the initial two paths between the α or β-secretases, depending on the first (also termed ε) cleavage by presenilin, CTFβ can either take the $A\beta_{40}$ (less pathogenic) or $A\beta_{42}$ (more pathogenic) product pathway[52]. The reasons for the variable activity of presenilin are not well understood, but what is clear is that mutations within the protease that favour $A\beta_{42}$ production are AD-associated[53]. Likewise, mutations within the APP gene that favour different cleavage patterns also lead to hereditary forms of AD, like the so-called Swedish mutation that creates a better substrate for the initial β-secretase, causing a significant increase in total Aβ production[54]. Additionally, there are also a wide range of mutations within Aβ that enhance its propensity to aggregate leading to hereditary forms of AD[50].

Regardless of how Aβ is produced, once it is allowed to accumulate to critical levels it will begin to oligomerise, initiating a complex cascade affecting many systems within the brain. Similar to other protein misfolding diseases, the oligomers of Aβ display toxicity and lead to significant levels of neuronal death themselves[55]. In particular, there is growing evidence that intracellular oligomer formation within the ER is particularly damaging to cells. This has been elegantly proven by Cattaneo and co-workers through the implementation of ER localised intrabodies that disrupt this pathway of Aβ aggregation[56,57]. They were able to demonstrate that these intrabodies ameliorated the defects of neurons isolated from Tg2576

transgenic mice, a common animal model of AD[57]. Besides their innate toxicity, it has become clear that Aβ oligomers, and other higher ordered species up to fibrils, cause further perturbations in the brain exacerbating the damage. Perhaps most significantly, these extracellular Aβ species lead to the onset of significant neuroinflammation through the brain's innate immune system. Beginning with microglia, the brain resident macrophage, Aβ oligomers have proven to induce a potent inflammatory response, releasing a plethora of cytokines that recruit and polarise other microglia into a neurotoxic state as well as astrocytes, glial cells responsible for maintaining the extracellular environment of the brain[58]. While there are many events that take place during Aβ mediated neuroinflammation, the most consequential appears to be the microglia/astrocyte mediated activation of the complement system (**Figure 1.5**). During development, the complement system is activated in the brain to prune excess synapses by microglial mediated phagocytosis and is necessary for proper maturation of the nervous system[59]. Hong and co-workers hypothesised that this system was being inappropriately activated during AD, leading to the synaptic loss that is strongly correlated with cognitive decline[60]. To investigate this, they studied the dynamics of synapse loss in J20 transgenic mice. This mouse strain has been developed to overexpress human APP with the aforementioned Swedish mutation, elevating the production of CTFβ, and the Indiana mutation, favouring the $Aß_{42}$ cleavage pathway by presenilin[61]. Using advanced super-resolution microscopy techniques, the team was able to show that in the synapses in the hippocampal and frontal cortex regions of AD mice were significantly more associated with the protein C1q, the initiating protein of the classical complement cascade (**Figure 1.5**). By using compounds that inhibit the γ-secretase complex, and therefore Aβ production, and studying the effects of localized injections of either Aβ oligomers or monomers into the brain, the researchers were able to show that synaptic targeting by C1q was enhanced by the presence of Aβ oligomers specifically. This complement mediated loss of synapses was also observed well before the deposition of plaques, an important observation from a therapeutic perspective since it supports the belief that early pre-plaque intervention is necessary to effectively combat AD[62]. Alongside other studies that have demonstrated that Aβ activated microglia further exacerbate the activation of the complement system by inducing astrocytes to produce the protein C3[63], the mediator of the alternative complement pathway, it is clear that on top of the inherent toxicity of Aβ aggregation, in AD Aβ has many avenues to initiate synaptic loss. However, the route to cell death, the hypothesis that elevated Aβ accumulation is ultimately the initiating factor leading the AD pathogenesis is becoming progressively more supported

despite doubts due to the failure of therapeutic interventions based on this hypothesis. Some of the reasons for these failures will be discussed in later sections.

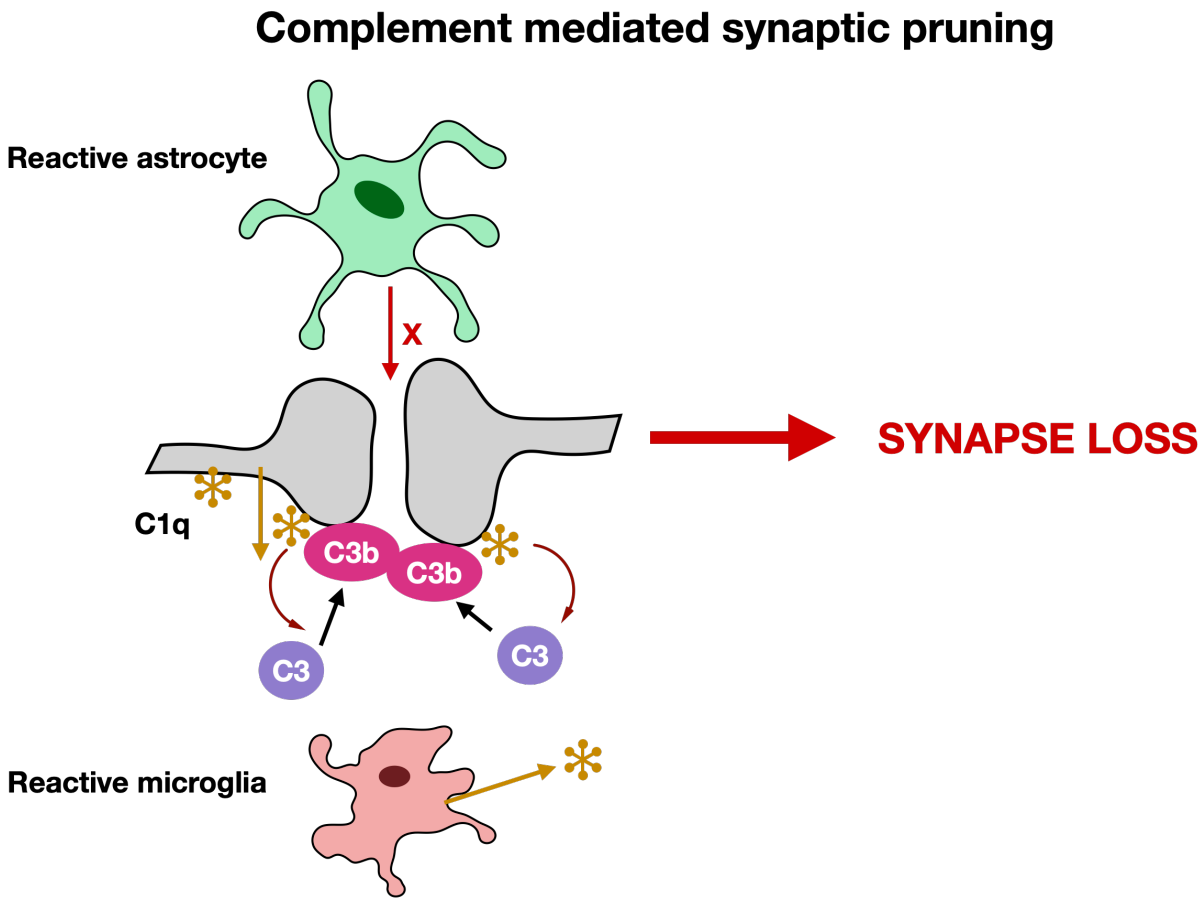


Figure 1.5. Complement system in AD. Illustration of the consequences arising from the mis-activation of the complement system in the brain during Alzheimer's disease triggered in response to Aβ oligomers. Through yet to be determined signals (represented by the red X), astrocytes can induce neurons to produce C1q themselves (yellow symbols). C1q is also created by reactive microglia. Opsonisation by C1q then encourages the cleavage of C3 into its active form C3b, which then further encourages synaptic phagocytosis through the alternative complement pathway.

1.1.4. Alzheimer's disease: Tau

Besides Aβ, the other major protein forming amyloid structures in AD is the protein tau. Unlike APP, tau has a well-known physiological role. Under normal conditions tau is predominantly expressed in neurons of the central nervous system, where it primarily stabilises neuronal microtubules (MTs) to promote axonal outgrowth[64]. Tau can be divided into three major domains, the C-terminal assembly domain responsible for MT binding, a proline-rich domain, and the N-terminal projection domain (**Figure 1.6a**). There are 6 main splice variants

of tau protein in the brain based on the inclusion or exclusion of different subdomains, all expressed from the *MAPT* gene, with the longest isoform being the 2N4R form at 441 amino acids. In this nomenclature, the 2N refers to the inclusion of two near-amino-terminal inserts and the 4R refers to the presence of 4 repeat domains within the MT binding domain, where only the second repeat is the varied domain (**Figure 1.6a**). Regardless of splicing, NMR and small-angle X-ray scattering studies have shown that tau is an intrinsically disordered protein in the absence of a binding partner[65]. Early studies of tau also found that it has a very low propensity to aggregate *in vitro* and established that it is an unusually hydrophilic protein with many polar or charged side-chains[64]. Therefore, the ability for tau to form well-ordered fibrillar structures *in vivo* appears to be a peculiar phenomenon. This peculiarity is further accentuated by the fact that fibrillar tau deposits are not only observed in AD, but in a whole family of neurodegenerative diseases termed tauopathies[66]. While some tauopathies have amyloid deposits consisting of all isoforms, such as AD, some diseases, like Pick's disease, only have a subset of the isoforms comprising the deposits[67]. From these observations, it is clear that some potentially very specific mechanisms regulating tau behaviour become dysfunctional in these disease processes.

Studies of the structures of tau deposits help to shed some light on how this seemingly stable protein can convert to the amyloid structure in so many diseases. Two hexa-peptide motifs within tau's repeat domain (^{275}VQIINK^{280} and ^{306}VQIVYK^{311}) have been shown to be crucial drivers of the amyloid formation of tau, despite the fact that they only represent ~3% of the total sequence (dependant on the isoform)[68]. Indeed, these sequences have the highest predicted β-sheet propensity and truncated forms of tau containing these motifs spontaneously aggregate[69]. Cryo-EM structures of the paired-helical filaments (PHFs) that compose NFTs also show that these structures are in the core of the fibrils, while the long N-terminus and shorter residual C-terminus of tau form a "fuzzy coat" surrounding this core (**Figure 1.6b**)[70]. What perturbations to tau could enable these motifs to drive amyloid formation? Another important characteristic of aggregated tau is that it contains many different post-translational modifications (PTMs). Most famously, tau in NFTs is found to be hyperphosphorylated. In the longest isoform, tau has a staggering 85 (80 Ser/Thr and 5 Tyr) potential phosphorylation sites. Unsurprisingly, phosphorylation is one of the primary mechanisms that regulates tau function *in vivo*. Normally, phosphorylation of tau is a key regulator for MT association, where the phosphorylation at key residues, in particular at various KXGS motifs, promotes the dissociation of tau[71]. When tau is bound to microtubules it does not fold into a consistent

globular structure, however, the aforementioned hexa-peptide motifs have been demonstrated to be absolutely essential to stabilising the interface of tau with the junction of α/β tubulin heterodimers, forming a local hairpin motif[72]. From these observations it is clear when tau is performing its predominant function, these potentially pathogenic motifs are sequestered from inter-tau association. Thus, processes that encourage the dissociation of tau from MTs add to the pool of tau available for aggregation. This highlights a larger theme within protein science of a competition between physiological function and pathological association[73].

While hyperphosphorylation is only one aspect of tau dysregulation within AD, it is an important process to understand, and recent studies have begun to elucidate the driving forces behind this dysfunction. Several genetic studies support that Aβ deposition is a prerequisite for AD-like tau hyperphosphorylation. Crosses of hAPP and hTau mice significantly increases tau deposition compared to the single transgene parents but has no effect on the levels of Aβ deposits[74]. Likewise, when cultured rat neurons expressing hAPP are exposed to Aβ oligomers isolated from AD patients, they exhibit neuritic dystrophy and tau hyperphosphorylation. However, if tau is knocked down before the exposure, no dystrophy is observed[75]. More mechanistic studies have revealed that Aβ oligomers can specifically activate mitogen-activated protein kinase (MAPK) and glycogen synthase kinase-3β (GSK3β), both of which have tau as a well-established target[76,77]. Greatly supporting this is a study by Zheng and co-workers using rat primary septal cultures showing that cultures exposed to Aβ oligomers and specific inhibitors of either MAPK or GSK3β attenuated the level of tau phosphorylation compared to cultures solely exposed to Aβ[78]. How Aβ drives phosphorylation exactly is still poorly understood, but it is clear that tau hyperphosphorylation can cause dysfunction by more means than just freeing up tau to aggregate. Hyperphosphorylation has been shown to drive the mislocalisation of tau from the axon to the dendrites, promoting neuronal dysfunction independent of aggregation[79]. Different phosphorylation patterns also alter the degradation routes of tau by inhibiting protease cleavage or the ability to associate with the proteasome[80,81]. Improper tau phosphorylation has even been shown to disrupt the kinesin complex, thereby interfering with global axonal transport[82]. It is humbling to be reminded that all of this complexity in disease progression mentioned, and much more not discussed here, is solely attributed to one type of PTM on tau. Indeed, hyperphosphorylation for many years has eclipsed the significance of the other PTMs that tau undergoes, but there are many more[83] (**Figure 1.6c**). It is clear that more methods and tools for accurately studying phosphorylation as well as the myriad of less well understood PTMs including acetylation, methylation,

glycosylation, ubiquitinoylation, and more need be developed to elucidate the importance of these modification within normal physiology and disease. Once developed, these tools could potentially drive the identification of new therapeutic avenues for intervention.

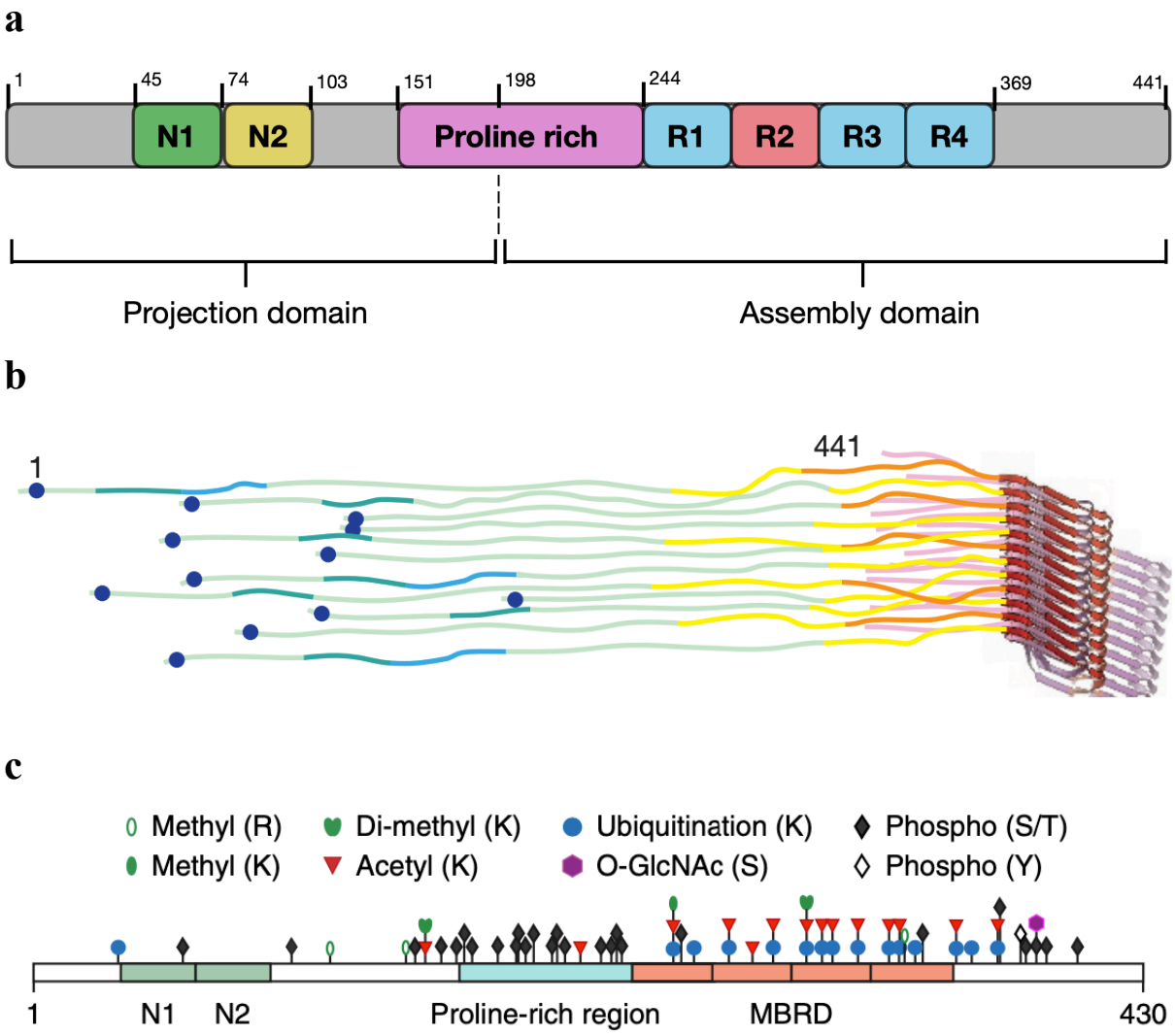


Figure 1.6. Tau structure and modifications. (a) Regions of tau, the variable repeat within the microtubule binding domain is (R2) is coloured red. (b) Illustration of the Cryo-EM structure of tau fibrils where the core (purple/red) region is formed from the R3 and R4 repeats and the "fuzzy coat" is comprised of the rest of the sequence. (c) Location and identity of the known PTMs of tau. Figures (b) and (c) adapted with permission from references [70] and [83], respectively.

1.1.5. Chemical kinetics for studying aggregation

What are the mechanisms of amyloid growth? How fast are the proteins associating? What microscopic steps can be inhibited? These are all important questions that require a level of detail beyond bulk macroscopic observations, and theoretical frameworks for describing the microscopic mechanisms behind amyloid formation are therefore needed. This type of demand

draws many parallels from the classical field of chemical reaction kinetics and the use of rate laws to elucidate reaction mechanisms. In this field, the usual workflow begins with the proposition of a differential rate law for the reaction being studied:

$$r = k[A]^a[B]^b \ldots \quad \text{(Eq. 1.1)}$$

Where [*A*], [*B*], and so on are the concentrations of the different species involved in the reaction, the exponents *a* and *b* are the reaction orders, and *k* is the rate constant. In order to make this equation convenient for comparing predictions to experimentally measurable outputs an integrated rate law must be derived, this allows the variation of the reactant concentrations in time to be measured and compared to the model. To illustrate, in the simplest case of *n*A → B the equations are as follows:

$$\frac{d[A]}{dt} = -k[A]^n \quad \text{(Eq. 1.2)}$$

$$\frac{1}{[A]^{n-1}} = \frac{1}{[A]_0^{n-1}} + (n-1)kt \quad \text{(Eq. 1.3)}$$

This last equation (1.3) represents the integrated rate law, where $[A]_0$ is the initial concentration. Now from this it is possible to derive that the half time of the reaction according to this model is:

$$t_{\frac{1}{2}} = \frac{\left(2^{n-1}-1\right)}{(n-1)k[A]_0^{n-1}} \quad \text{(Eq. 1.4)}$$

A key observation from the above equation (1.4) is that the reaction half time should scale with the initial reaction concentration according to $t_{1/2} \propto [A]_0^{\gamma}$, where $\gamma = -(n-1)$. This scaling component is directly related to the reaction order, or the number of microscopically interacting species. So, through this simple set of operations a now easily measurable macroscopic output, reaction halftime, can elucidate the microscopic mechanisms underlying the reaction. This powerful approach has been used innumerable times in physical organic chemistry, and a similar framework for the study of protein aggregation has evolved directly from these principles.

To begin to apply this methodology to protein aggregation, it helps to first think of the possible mechanisms that lead to an increase of fibril mass and number (**Figure 1.7a**). At a high level, there are two main classes of mechanisms: nucleation mechanisms that create new fibrils (primary nucleation, surface catalysed secondary nucleation, and fibril fragmentation) and growth mechanisms that alter the mass of existing fibrils (elongation and dissociation at the fibril ends). In the absence of any preformed fibrils, primary nucleation where monomers

associate to form competent nuclei is always the first event, it is therefore considered a primary pathway within the nucleation mechanisms. However, primary nucleation can occur through homogenous or heterogeneous mechanisms. In a heterogenous pathway, specific interfaces promote the formation of the growth competent nuclei, such as in the case of cholesterol for Aβ[84]. While it may seem similar, heterogenous nucleation is distinct from the fibril surface catalysed secondary nucleation. This is because the fibril amount is varying throughout the course of the reaction, and thus the catalytic contribution is variable while it is fixed for heterogenous nucleation. Along with fibril fragmentation, surface catalysed nucleation falls into the secondary nucleation pathways that are dependent on fibril mass. These secondary mechanisms can also be thought of as autocatalytic feedback loops as they promote the exponential increase of aggregate number through their self-replication. For the growth mechanisms, monomer dissociation from the fibrils has been shown to be much slower than association and can be neglected[85]. With this picture of the different mechanisms underlying amyloid formation we can now attempt to place these into a differential rate law that describes the population balance of each of the relevant species that are present throughout the reaction, otherwise known as a master equation[86]. The derivation of the master equation for amyloid growth is beyond the scope of this thesis, but the importance of its application to global fitting of aggregation curves to elucidate mechanisms is hard to overstate. When coupled with procedures for reproducible aggregation reactions, which are monitored by the fluorescence of the dye thioflavin-T (ThT) that fluoresces upon intercalating in the β-sheets of amyloid fibrils, the exact microscopic mechanisms can be reliably determined. Properly fitting experimental aggregation reactions with this method has established the mechanisms for many disease associated proteins. A well-established example is how this method elucidated that the proliferation of Aβ is heavily governed by surface catalysed secondary nucleation, shedding light on the mechanism behind decades of clinical observations of “halos” of oligomeric aggregates surrounding mature amyloid plaques in patient brains[87].

The utility of these equations extends further to studying the specific perturbations of inhibitors of aggregation as well, providing much needed guidance in drug development. This framework is again analogous to classical drug discovery for enzyme inhibitors, where the Michaelis-Menten equation is applied to fit experimental data in order to determine the mechanism of inhibition (i.e. competitive, non-competitive, etc.). Similar to this, the key approach to applying chemical kinetics to studying aggregation inhibitors is to realise that alterations to different microscopic events result in different macroscopic aggregation profiles

(**Figure 1.7b**)[88]. Alterations to the rates within the master equation can then be explored to find best match the experimental data, and thus determine the microscopic steps influenced by the inhibitor, which are a direct reflection of the potential species the inhibitor interacts with. This type of analysis is crucial for evaluating anti-aggregation-based therapeutics, as the haphazard inhibition of fibrillization can actually increase the production of toxic oligomers[86]. This type of approach has already facilitated the rational generation of several small molecule inhibitors of Aβ and α-synuclein aggregation[89,90].

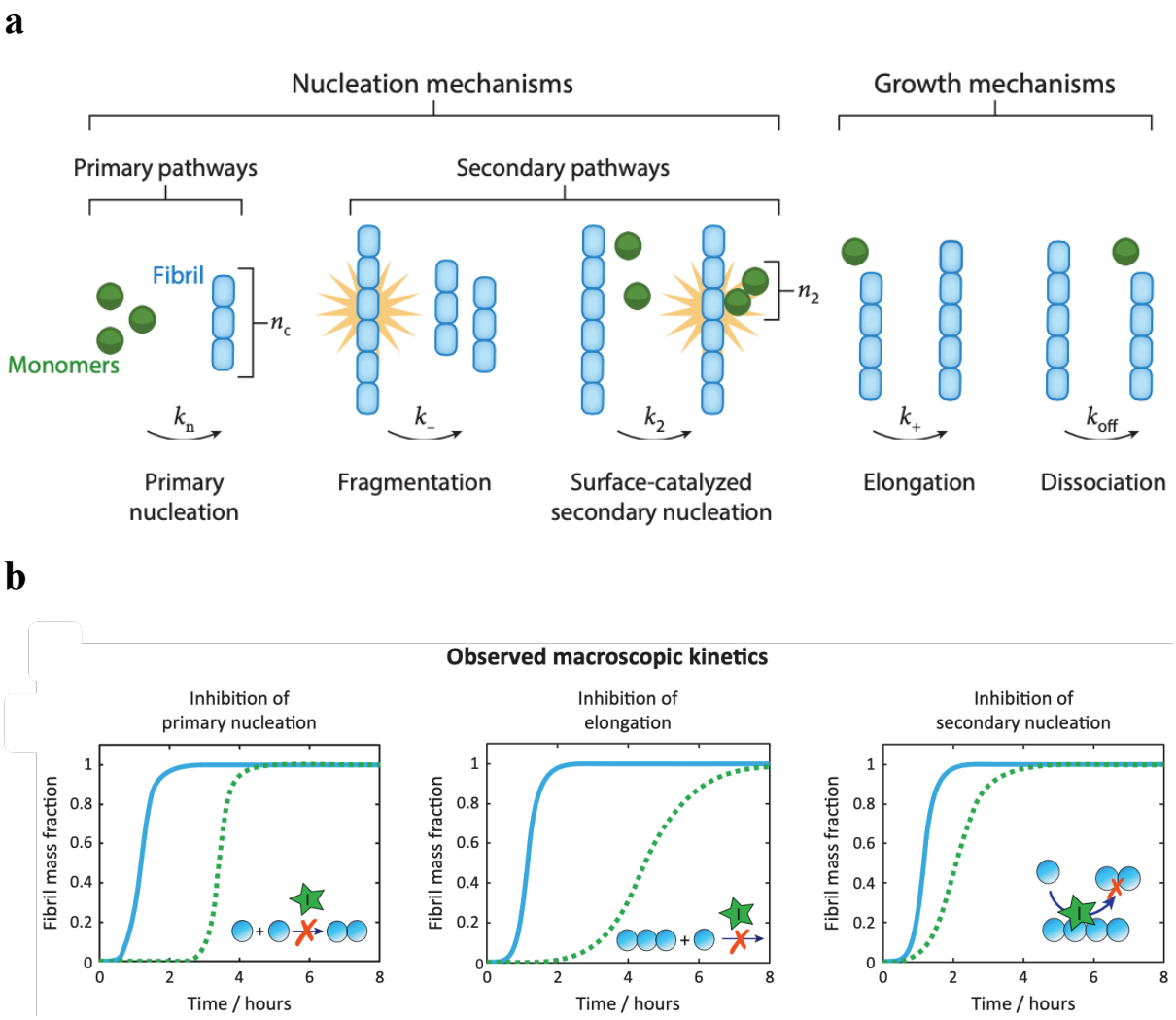


Figure 1.7. Chemical kinetics of protein aggregation. (a) Illustration of the aggregation mechanisms that increase fibril number and/or mass. (b) Macroscopic perturbations of the aggregation profile brought about by various modes of inhibition. Figures adapted with permission from references [85] and [88].

1.2 Antibodies

1.2.1. Antibody structure and diversity

Antibodies have become a cornerstone in medicine for both therapeutic and diagnostic purposes and have become invaluable tools for basic research. Nearly all of the applications of antibodies derive from their incredible ability to bind to virtually any desired target with high affinity and specificity. The scope of applications of this fundamental capability is remarkable, from cancer therapy[91], to the selective delivery of cytotoxic agents[92], to enabling super resolution microscopy[93], to medical imaging[94], and even to act as catalysts[95]. These developments are reflected in the fact that the global market for monoclonal antibodies (mAbs) exceeded $100 billion USD in 2019.

The overall structure of antibodies is highly conserved between species. While there are several isotypes of antibodies in humans, the vast majority of antibody-based immunity is conducted by the canonical immunoglobulin G (IgG) type (which itself has four subclasses). IgG's are large (~150 kDa) and are composed of four peptide chains linked together by disulphide bonds, two larger heavy-chains and two smaller light-chains forming an overall "Y" shape (**Figure 1.8a**)[96]. The stem of the "Y" is formed by the C-terminal regions of the heavy-chains and is termed the fraction-crystallizable (Fc) and is responsible for signalling with other entities in the immune system through highly conserved N-glycosylation sites. Each arm of the "Y" is identical and is termed the antigen-binding fragment (Fab), meaning a full IgG contains two antigen binding sites. The Fab region can itself be differentiated into one constant and one variable domain, each consisting of segments from the light and heavy-chains. The variable regions (Fv) of the heavy and light chains (V_H and V_L, respectively) contain the antigen binding sites. Binding is specifically mediated through the 3-complementary determining region (CDR) loops on each variable fragment (**Figure 1.8b**). The sequence and size of these loops varies wildly between antibodies, understandably since these are the regions responsible for recognizing the particular target the antibody is raised against.

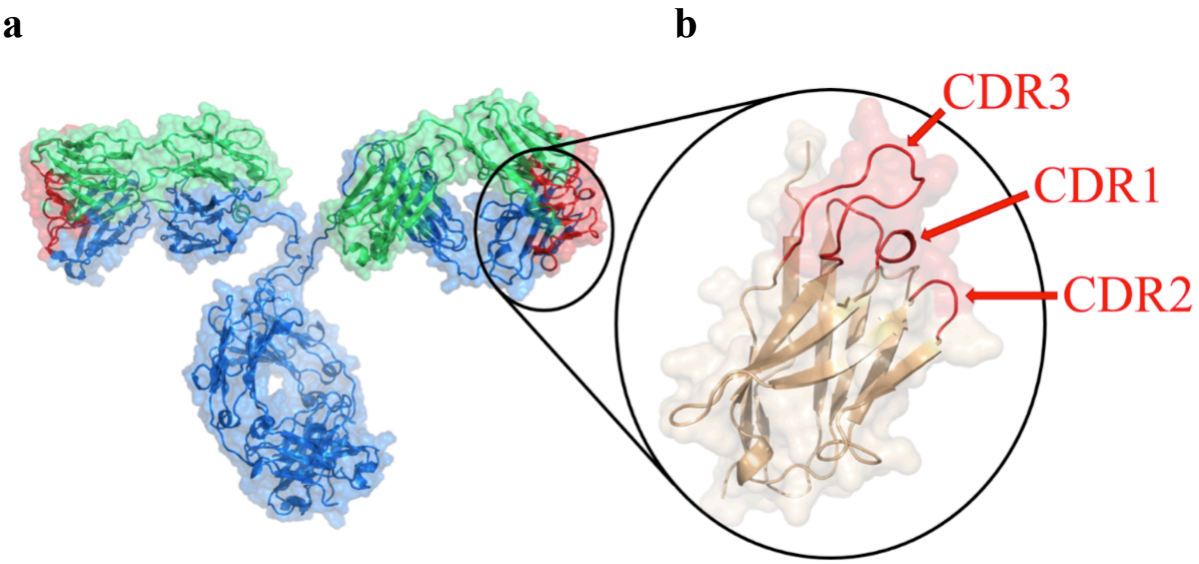


Figure 1.8. Antibody structure. (a) Structure of a full-length IgG, heavy chains are shown in blue and light chains in green with the CDRs of each in red (PDB: 1IGT). (b) Close up of the CDR loops of the variable domain of the heavy chain with each of the three loops indicated.

There are a few notable exceptions to the general IgG structure. The sera of the Camelidae family contain very unusual heavy-chain antibodies (HCAbs) that are completely devoid of light-chains and lack the first constant domain[97] (**Figure 1.9a**). Quite remarkably, similar immunoglobulins have also been found in certain species of shark[98]. There are several subtypes of HCAbs found in camelids with different hinge region lengths. The minimal binding entity that conventional heterotetrameric IgGs can be reduced to is the tethered V_L and V_H fragments (scFv). However, since HCAbs lack a light chain, they can be reduced to a single variable heavy-chain fragment (VHH), termed a nanobody (Nb) (~14 kDa depending on the sequence) (**Figure 1.9a**). VHHs differ from the V_H of normal IgGs in a few significant ways that enable them to function in the absence of a light-chain. Conventional V_H's have several highly conserved hydrophobic amino acids at the interface with the V_L, in camelid VHHs these are mutated to smaller or hydrophilic amino acids to create a less "sticky" surface that can be stable in the absence of a binding partner[99]. These structural adaptations have now been incorporated into human VHH's in order to create stable fully human single domain antibodies (sdAbs). Another consequence of removing the light chains is that antigens can only be recognized by 3 CDRs instead of 6. VHHs overcome this loss of interacting surface by extending the CDR loops, the CDR3 loop most significantly[100] (**Figure 1.9B**). Their reduced size and extended loops have famously enabled sdAbs to access "cryptic" epitopes buried in proteins previously inaccessible with conventional antibodies[101]. Reducing an antibody down

to its minimal binding element also allows for specific binding to the desired target but without eliciting the immune response, which again is mediated through the Fc region. This is highly desirable when inflammation should to be avoided, such as with neurological diseases. Additionally, the reduced size of sdAbs makes them more efficient at tissue penetration and even aids them in crossing the blood-brain barrier[102].

However, there are certain physiological drawbacks with the reduced size and complexity of sdAbs. The drawbacks are not only limited to the avidity, as their small size is also significantly lower than the renal threshold of 60 kDa causing their serum half-life to be significantly shortened. Their lack of an Fc region means they cannot engage the neonatal Fc receptor (FcRn) as well, which normally protects IgGs from lysosomal degradation[103]. Thus, this further reduces their serum half-life.

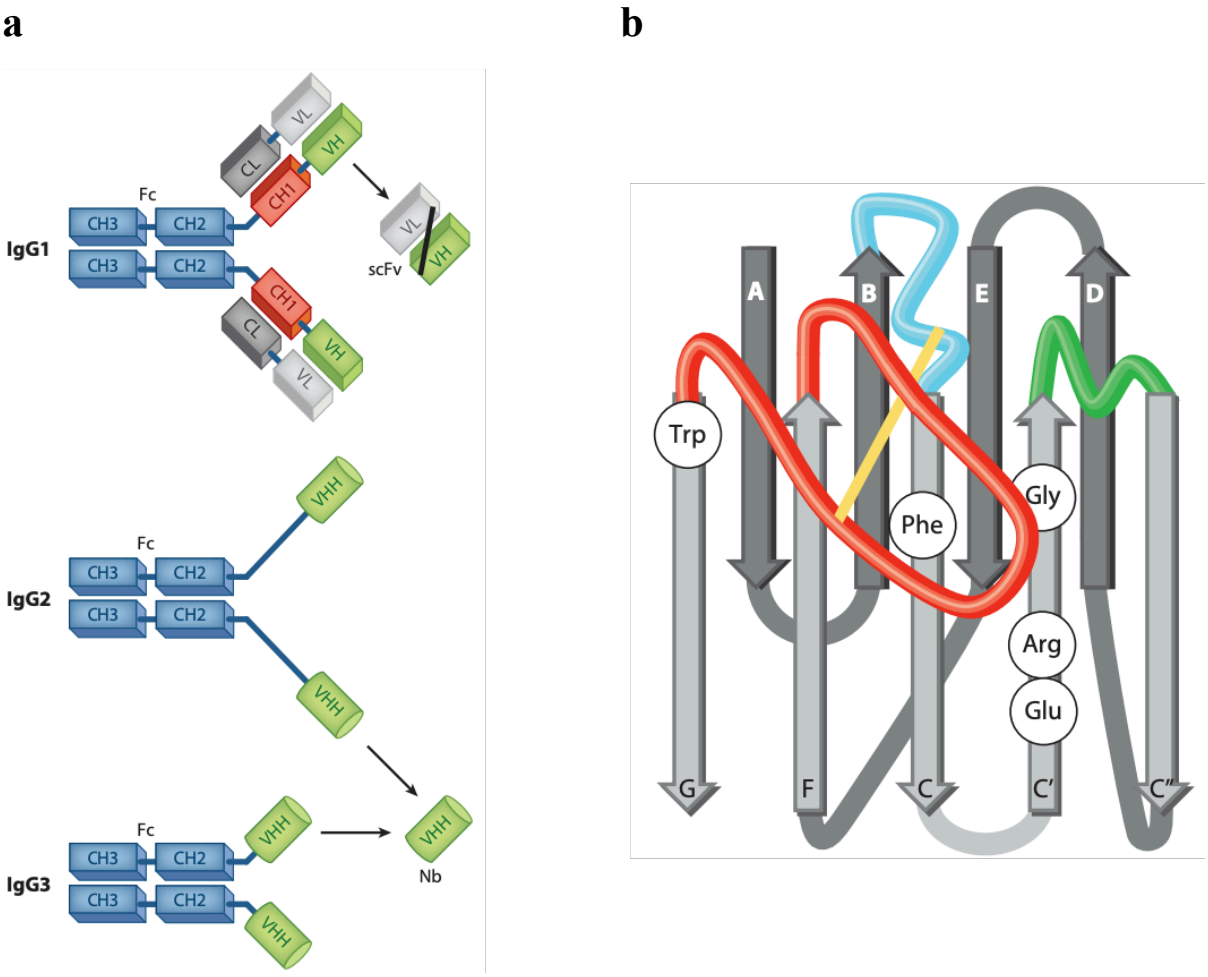


Figure 1.9 Camelid antibodies and nanobody structure. (a) Schematic representation of the different types of antibodies found in camelid serum and their minimal binding domains. (b) Structure of a canonical camelid nanobody with the extended CDR3 loop in red, CDR2 in green and CDR1 in blue. Key amino acid positions that would interact with a variable light-chain are shown in their 3-letter code, these have been mutated from less favourable amino acids to increase stability when exposed to the solvent or are covered by the extended CDR3 loop. Figures adapted with permission from reference [97].

1.2.2. Antibody maturation *in vivo* and endogenous functions

Within normal physiology, antibodies are generated through a series of processes in the immune system with the ultimate purpose of binding foreign antigens to mediate the host defence response in adaptive immunity towards pathogens. New antibodies are generated through a short lived but rapid evolutionary process that is confined within an individual's immune system[104] (**Figure 1.10**). The development begins with the capture and presentation of a foreign antigen by follicular dendric cells (FDCs) inside compartments called germinal centres. In these centres, B-cells present highly diverse antibody paratopes that have been generated through V(D)J recombination and somatic-hypermutation (SHM), which creates a large amount of diversity in the CDR loops by a localised error-prone DNA replication process. These paratope presenting B-cells are then subjected to a competition for binding the antigen presenting FDCs, where unbound B-cells with poorly performing paratopes are removed through apoptosis, thus refining the library. The bound B-cells are then proliferated and undergo further rounds of SHM and selection, successively increasing their binding affinity and specificity. Ultimately, when the binding affinity is optimal, the B-cells are transformed into memory B-cells that now can rapidly generate their refined antibody in response to re-exposure of the foreign particle[105].

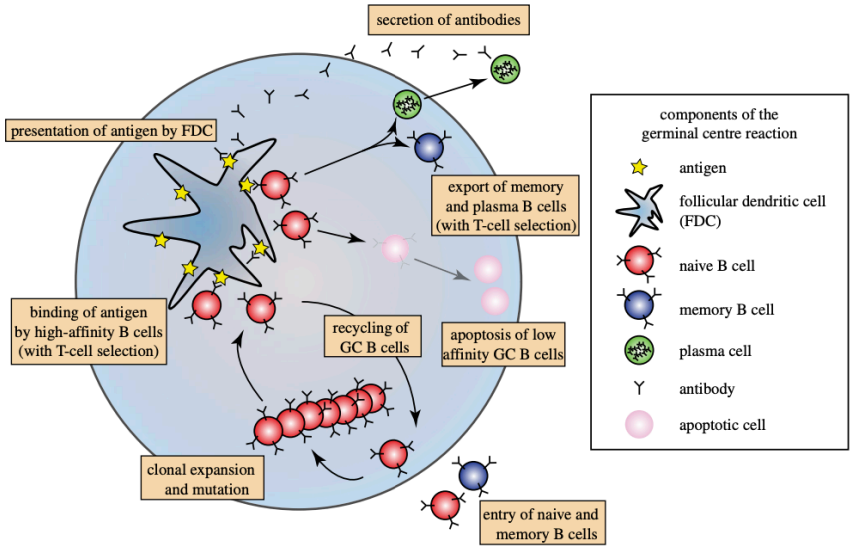


Figure 1.10. *In vivo* antibody maturation. Illustration of the process and components responsible for the somatic hypermutation leading to mature antibodies for adaptive immunity. Figure adapted with permission from reference [105].

Target engagement by an antibody can lead to several outcomes. As mentioned previously, the Fc region of the antibody is responsible for signalling to immune cells within the body through Fc receptors. When an antibody is bound to a target, it can be marked for clearance through phagocytosis by bringing together the target and phagocytes through this dual interaction in a process called antibody opsonisation[106] (**Figure 1.11a**). This type of activity is regulated by the patterns of glycosylation on the Fc region of the antibody. Antibodies are also capable of mounting a defence solely through their own action, such as by blocking the activity of a critical region of the pathogen crucial for biological activity, such as cellular entry receptors of a virus[107] (**Figure 1.11b**). The multi-avidity of antibodies enables yet another mechanism of protection through a network of interactions between several antibodies and their targets in a process called agglutination[108]. This mechanism limits the spread of the pathogen or even precipitates them out of solution and localises them for efficient clearance by other immune cells (**Figure 1.11c**).

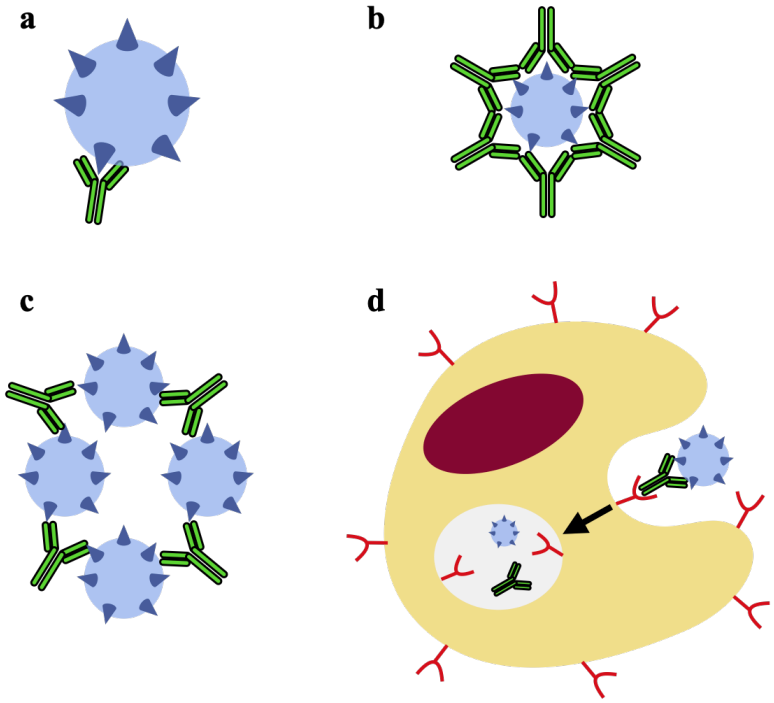


Figure 1.11. Modes of antibody function. (a) Opsonisation of a pathogen (blue) occurs when an antibody (green) marks the object for clearance through binding its antigen. (b) A pathogen can also be neutralised through antibodies completely coating the part of entity crucial for pathogenicity. (c) The multi-avidity of antibodies can also lead to the agglutination of many pathogens that can cause them to precipitate from solution and thus loose pathogenicity. (d) Antibody mediated phagocytosis is triggered when Fc receptors (red) bind to an engaged antibody, stimulating the phagocyte (yellow) to endocytose and destroy the material.

1.2.3. Classical directed antibody maturation

For the last several decades, antibody maturation in biopharma has been driven by attempting to mimic the rapid evolutionary process utilised by nature. A number of techniques to accomplish this have been developed and are collectively called directed evolution. While directed evolution is being applied in ever increasingly sophisticated ways for complex functions, for antibody binding the techniques are relatively straightforward and usually fall into the category of display methods[109]. Regardless of the approach, all directed evolution projects begin with the diversification of the starting protein of interest's DNA. The starting protein is usually a naturally occurring protein with some of the already desired characteristics or activity[109]. The diversification can be achieved through multiple approaches such as error-prone PCR or saturation site-directed mutagenesis. The power of directed evolution techniques then comes through the linkage of the coding sequence to the physical activity of its product. For display methods this can be achieved through multiple approaches including phage display, ribosome display, yeast display, and mRNA display (**Figure 1.12a-c**). While the vehicle for the protein and its encoding DNA/RNA is different in each of these, the principles are largely the same. The diversified library is exposed to the immobilised desired target and undergoes competitive binding, then poorly bound or unbound candidates are washed away from the displayed antigen. After several rounds of selection high-affinity, antibodies are isolated and sequenced. The selection of the platform depends on the size of the protein to be developed; yeast-surface display is capable of evolving full IgGs thanks to the eukaryotic quality control machineries yeast contain to assist with protein folding. However, this comes at a cost of the size of the library that can be screened (~10^7 variants), largely limited by the transformation efficiencies of the yeast themselves[110]. Following this trend, ribosome and mRNA display can have libraries as large as 10^{13} variants since they are entirely *in vitro* and therefore not limited by transformation, but due to the lack of quality control mechanisms these are best suited for the creation of smaller polypeptides[111].

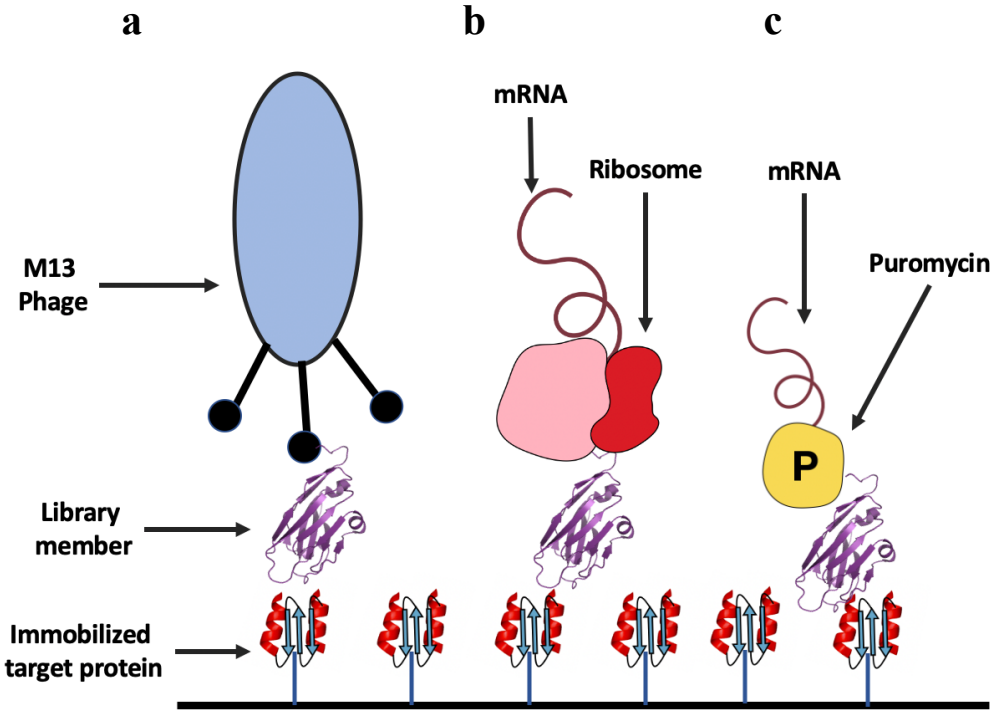


Figure 1.12. Examples of *in vitro* display methods. (a) Phage display relies on *E. coli* to synthesise and fold proteins that are fused to the coat proteins of viruses. (b) Ribosome display takes advantage of conditions to prevent release of the protein and mRNA (c) mRNA display utilises puromycin to covalently link a protein with its mRNA after translation and release from the ribosome.

1.2.4. *In silico* antibody development

The combination of *in vivo* and *in vitro* approaches to antibody development have enabled the great success of antibody-based therapies that we have seen in the last 25 years. However, these approaches still face limitations in a number of areas within the development process[112]. Several target classes remain as difficult targets for discovery programs, including membrane proteins, aggregation-prone targets, selection of agonists versus antagonists of biological pathways, and highly homologous targets. While both the classical immunization and *in vitro* approaches for antibody development offer a vast repertoire of unique sequences, these techniques primarily select for simply the tightest binders of the target, which usually occurs at immunodominant epitopes that may not be the functionally relevant sites[112]. If there are lower affinity binders at the desired sites, these are not likely to be selected through these methods. This critical fault has driven the development of advanced experimental apparatuses to enable selection based on functional screening rather than binding. Nonetheless, these approaches are costly and may not be applicable in every case[113]. Ultimately, this has led to

the development of a number of *in silico* approaches to design antibody paratopes to target weakly immunogenic sequences or specific functional sub-domains. The successful development of approaches in this area of research could make massive strides in streamlining and reducing the cost of antibody drug discovery.

Intrinsically disordered proteins (IDPs), a major class of proteins involved in a wide range of biochemical processes, and the class that many amyloidogenic proteins belong to, can pose as difficult targets for classic antibody development techniques[114]. This is partly due to their largely linear nature which makes it difficult to isolate antibodies for specific functional or pathological sub-domains. To address this, a computational method has recently been developed to generate paratope sequences against linear targets[115]. The method is centred on a cascade approach for the generation of complementary sequences derived from the analysis of structures in the Protein Data Bank (PDB). Once a query target sequence is chosen, the PDB is scoured for structures where pieces of that sequence, of three or more residues, is engaged in a face-to-face β-strand, either parallel or antiparallel. Hopefully, the search identifies a series of fragments (β-pairs) that overlap to cover the entire queried sequence, this library of complementary sequences can then be assessed for their compatibility for merging into a peptide that can cover the entirety of the target. In brief, to accomplish this first the β-pairs must be of the same type of (i.e. both antiparallel), they must contain at least two overlapping residues at the merging junction, and the hydrogen bonding pattern of the backbone of overlapping region must match. In the end, the principle of this method is that, once these complementary sequences are merged, the ability of each fragment to interact with the target in a β-sheet is additive and will contribute towards a larger overall β-sheet interaction with the entire linear target (**Figure 1.13a**). This sequence can then be grafted into the CDR loops of an antibody to confer binding, and has been successfully applied to develop single-domain antibodies against α-synuclein as well as a panel against the sequence of Aβ and has produced significant data to support the binding site is indeed for the desired location through experiments like the loss of binding after proline insertion into the epitope (**Figure 1.13b,c**).

When targeting aggregation prone regions of proteins, there is a high risk of producing poorly soluble paratope sequences within the antibody[116,117]. To avoid this, solubility prediction algorithms can be run in parallel to evaluate engineered sequences[116]. From an industrial perspective, this is vital as many candidate antibodies that display desired activity may still be unfeasible for production if their solubility limits formulation strategies or if their instability is too great[119]. An example of a solubility prediction algorithm is CamSol, and its

application in tandem with rational paratope design represents what is emerging as the third generation of discovery methods for antibodies (**Figure 1.14d**).

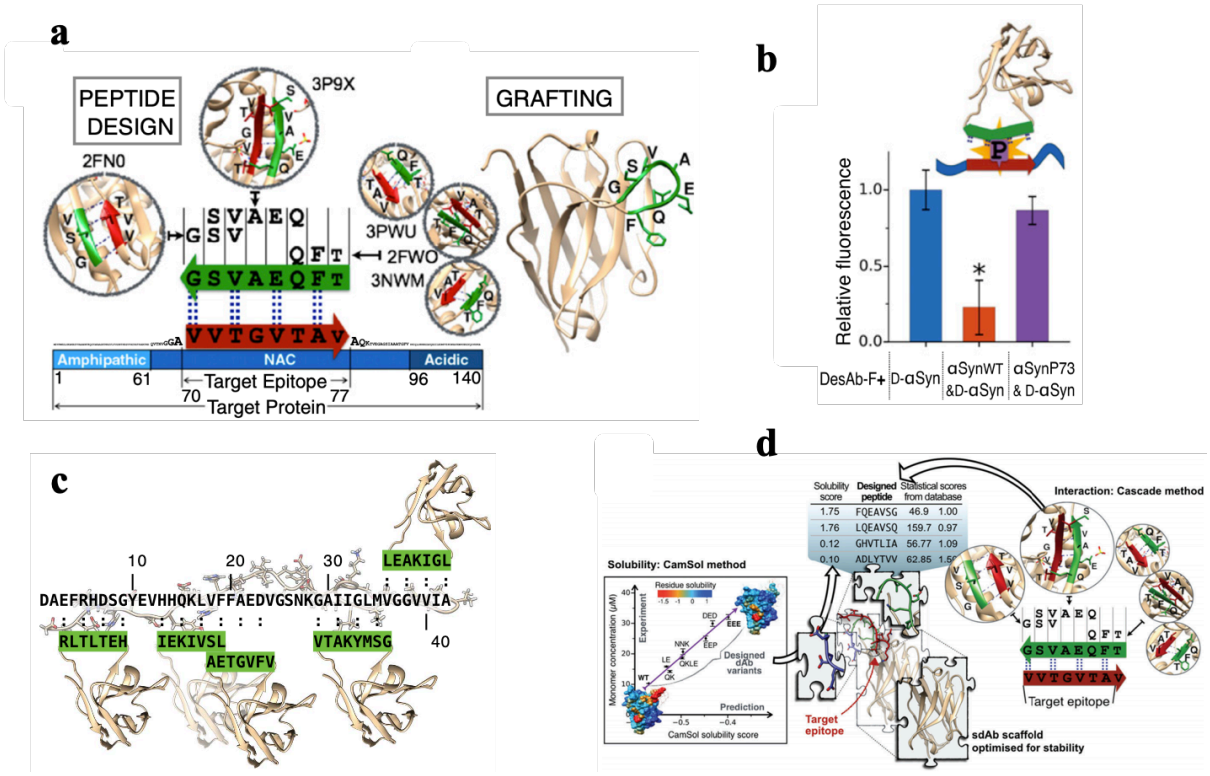


Figure 1.13. Rational design of antibodies. (a) Schematic of the creation of a complementary peptide graft for the NAC region of α-synuclein. Three different properly overlapping sequences found in the PDB engaging the desired sequence in a β-strand conformation are combined to create the desired interaction. (b) A Fluorescence competition assay of the designed single domain antibody for labelled α-synuclein, blue is in the absence of non-labelled competitor, red is in the presence of equimolar unlabelled α-synuclein and purple is in the presence of equimolar unlabelled α-synuclein with a proline inserted in the target sequence (P73). (c) Panel of single-domain antibodies designed to span the length of Aβ. (d) Illustration of various rational design strategies to improve antibody function, stability, and solubility. Figures adapted with permission from references [115] and [112] respectively.

1.2.5. Antibody use in Alzheimer's disease

Antibodies and fragments thereof have been developed or discovered against a wide range of protein states beyond the native state, including transition and aggregated states[112]. In the case of Aβ, Glabe and co-workers have leveraged classical immunization techniques by injecting rabbits with stabilised $A\beta_{40}$ oligomers covalently linked to gold nanoparticles to obtain a panel of oligomer specific mAbs[120]. The variable reactivity of this panel revealed that amyloid oligomers are structurally diverse and at least a subset can propagate their structures

in a prion like fashion. Prior to this study, Cattaneo and co-workers had developed a panel of intracellularly stable oligomer specific scFvs through a more complex intracellular selection platform[121]. The platform is based on screening scFv cDNA libraries in yeast derived from naïve and Aβ-immunised mice for their ability to associate with an Aβ-bait that initiates the translation of an essential gene. This study yielded unique scFvs that displayed both linear sequence specificity and conformational sensitivity likely arising from the more physiologically relevant conformations and the possible oligomers that the Aβ-bait adopts in the intracellular environment compared to other campaigns that utilise immobilised Aβ[122].

The power of antibodies to treat infectious diseases and cancer similarly led to the investigation of their ameliorative ability in AD beginning in the late 1990's. Active and passive immunization strategies have both been explored clinically, although passive immunizations have become the much more dominant approach after an unfortunate early trial of an active immunization observed the development of T cell mediated meningoencephalitis in a significant number of patients[123]. For passive strategies, the dominant target for the last two decades has been Aβ, based on the amyloid hypothesis of AD (although several novel therapies based on other targets in the disease cascade have begun being explored more recently)[124]. A plethora of antibodies, all full-length IgG mAbs, against various forms of Aβ have been developed by a handful of pharmaceutical companies (**Table 1.2**), the first being bapineuzumab from Pfizer[125]. Famously, all but one of these candidates has failed to show significant efficacy in phase II or III of clinical trials, where the primary measures of efficacy are cognitive improvements and the reduction of fibrillar Aβ in the brain as measured by positron emission tomography (PET) scans. The common causes cited for the failure of these trails have been either administrating too late in the pathology of the disease, therefore the activity of these agents is too little too late, or that the dosing was insufficient[126]. From a practical perspective, the latter issue is one that might potentially be more tractable. This is because most of the mAbs against Aβ have exhibited well tolerated safety at the doses that have been explored[127]. The former issue is of a somewhat more convoluted nature since two perspectives can be taken: either early intervention is critical for any treatment benefit, or in late disease stages with advanced Aβ deposition modest clearance is simply insufficient or even irrelevant[58]. Despite the unresolved ambivalence, current trials are unquestionably aiming to treat patients in earlier disease stages.

Antibody	Manufacturer	Origin	Subclass	Epitope
Bapineuzumab	Pfizer/Janssen	Humanized	IgG1	AA 1-5
Solanezumab	Eli Lilly	Humanized	IgG1	AA 16-26
Gantenerumab	Hoffman-La Roche	Human	IgG1	AA 3-12, AA 18-27
Crenezumab	Genentech Inc.	Humanized	IgG4	AA 13-24
Ponezumab	Pfizer	Humanized	IgG2	AA 30-40
BAN2401	BioArctic, AB/Esai Co.	Humanized	IgG1	Protofibrils
Aducanumab	Biogen Inc.	Human	IgG1	AA 3-6 (conformation specific)

Table 1.2- Clinical monoclonal antibodies for Aβ

The antibody that is the closest to Food and Drug Administration (FDA) approval for AD treatment is aducanumab by Biogen, which is currently in an extended phase III trial. However, the trial was only extended after significant post hoc analysis after it was initially discontinued by an earlier assessment claiming it was unlikely to meet its target endpoint[128]. As seen in Table 1.2, aducanumab targets the N-terminus of Aβ (residues 3-6), but only when it is in a conformational epitope found in oligomers and fibrils[129]. To attempt to rationalise why aducanumab achieved relative success compared to its predecessors, a kinetic study of the aggregation inhibition of a panel of clinical Aβ antibodies was recently conducted[130]. This illustrated that aducanumab's binding mechanism is the most efficient at reducing the oligomeric population during the reaction progression, perhaps explaining why aducanumab also showed the highest reduction of amyloid plaques in the clinic as measured by PET (**Figure 1.14a,b**). These findings indicate that previous attempts to generate clinical Aβ antibodies were likely generated by a process that does not fully appreciate the complexity of the aggregation reaction. If aducanumab does achieve FDA approval, it is by no means the end of the race for a disease modifying AD therapy; as mentioned above, a post hoc analysis was needed to justify the continuation of aducanumab's trial. This analysis revealed that only patients in the early

stage of the disease receiving the highest dose exhibited any clinical benefit[128]. This means that patients in the midst of the disease likely still need alternative therapeutic approaches.

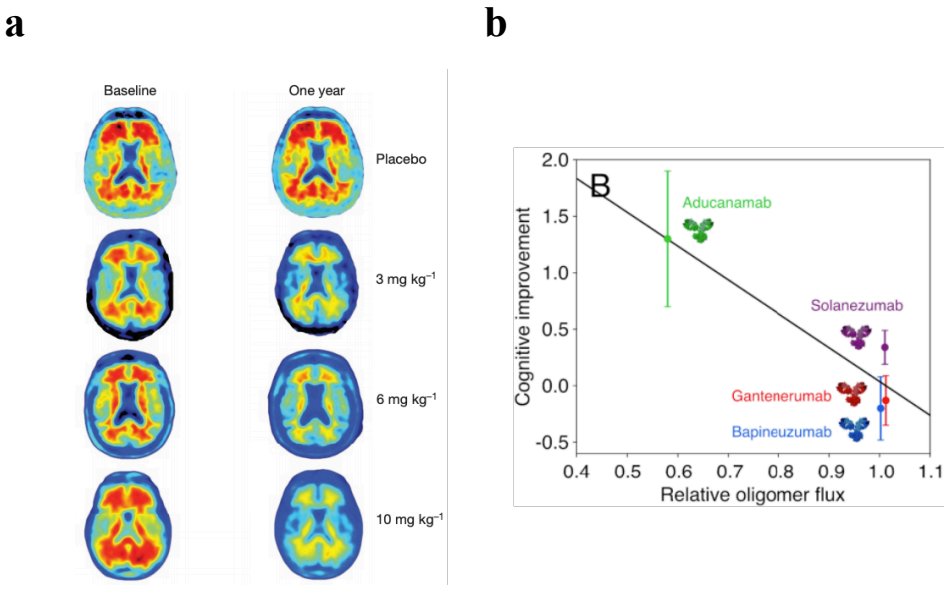


Figure 1.14. Aducanumab *in vivo* and *in vitro* activity. (a) Axial slices of PET images showing amyloid burden in the posterior of patient brains on various dosages of Aducanumab after 1 year of treatment. (b) Graph displaying the correlation of the measured changes in oligomer flux of the aggregation reaction in the presence of various clinical antibodies with their reported cognitive improvement. Figures adapted with permission from references [129] and [130] respectively.

1.3 Bioorthogonal chemistry

1.3.1. Chemistry in living systems and bioorthogonal chemistry

Life has evolved to perform chemical transformations under a relatively narrow range of conditions. Biological systems operate exclusively between 0 and 100 °C, in aqueous solution, with limited functional groups and metal ions, and in highly dense and heterogeneous environments. To perform the chemical transformations life requires under these conditions, nature has evolved enzymes as incredible biological catalysts and has created specialised organelles with optimised conditions for particular reactions, amongst many more adaptations. Since the dawn of biochemistry, scientists have tried to develop approaches to subtly manipulate and probe biological processes within these confined conditions. Great advances in this realm were enabled through the advent of fluorescent proteins and efficient genetic engineering tools, which enabled proteins of interest to be selectively labelled at their synthesis for monitoring[131]. However, there are a number of limitations with genetic labelling strategies. These approaches lack temporal control, cannot label other biomolecules such as lipids or

nucleic acids, cannot study post-translational modifications, and the large protein tags can also interfere with the endogenous function of the protein of interest[132]. These limitations, amongst others, have prompted chemists to work alongside biologists to develop reactions that can selectively target unique functional groups while not interfering with other biological processes, otherwise known as bioorthogonal chemistry.

For a reaction to be considered useful from a bioorthogonal perspective, it needs to meet some specific criteria: 1) it needs to be compatible with aqueous media, 2) at physiological pH, 3) have fast reaction rates at physiological temperatures and low reactant concentrations, and 4) it needs to be inert to the functional groups in all the various biologically active molecules[133]. From the perspective of *in vivo* protein modification, none of the canonical amino acids can function as a unique moiety for tagging. To overcome this, early examples of bioorthogonal reactions relied on uniquely reactive peptide motifs composed of natural sidechains but in an uncommon sequence. An early example of this is the tetracysteine motif used for selectively labelling proteins with membrane permeable fluorescent biarsenical compounds like FlAsH and ReAsH[134] (**Figure 1.15a**). Other approaches similar to this rely on peptide sequences capable of coordinating functionalised metal ions like histidine rich sequences and Ni-NTA probes as well as the more nuanced use of trans-splicing inteins[135,136]. However, with using natural functional groups there is a very limited diversity of approaches for *in vivo* chemical labelling.

While less complex than labelling inside the cellular environment, several aspects of selective protein modification *in vitro* with purified proteins still remain challenging. The purpose of *in vitro* labelling can go far beyond imaging aides, such as the attachment of cytotoxic reagents in the case of antibody-drug conjugates[137], adding poly-ethylene glycol (PEG) chains to improve the serum half-life of a protein,[138] or studying difficult to access PTMs and much more[139]. The nature of these applications places great weight on the homogeneity of the conjugation reaction, the stability of the conjugate, and the versatility of the chemistry. *In vitro* labelling has typically relied on the unique reactivities of specific canonical amino acids, most famously the nucleophilicity of lysine and cysteine residues (**Figure 1.15b**). However, traditional conjugation approaches at these and other residues have long had difficulties with meeting some of the above criteria, such as the challenges with NHS esters and malimides[140]. This has led to the exploration of more advanced conjugation reagents, as well as the investigation of the introduction of more appropriate amino-acid functional groups into proteins. The latter approach is attractive for a number of reasons, as inserting a new non-canonical amino acid can enable huge advancements for both *in vivo* and *in vitro* conjugation

issues, and the research in this area has been experiencing rapid development over the last two decades.

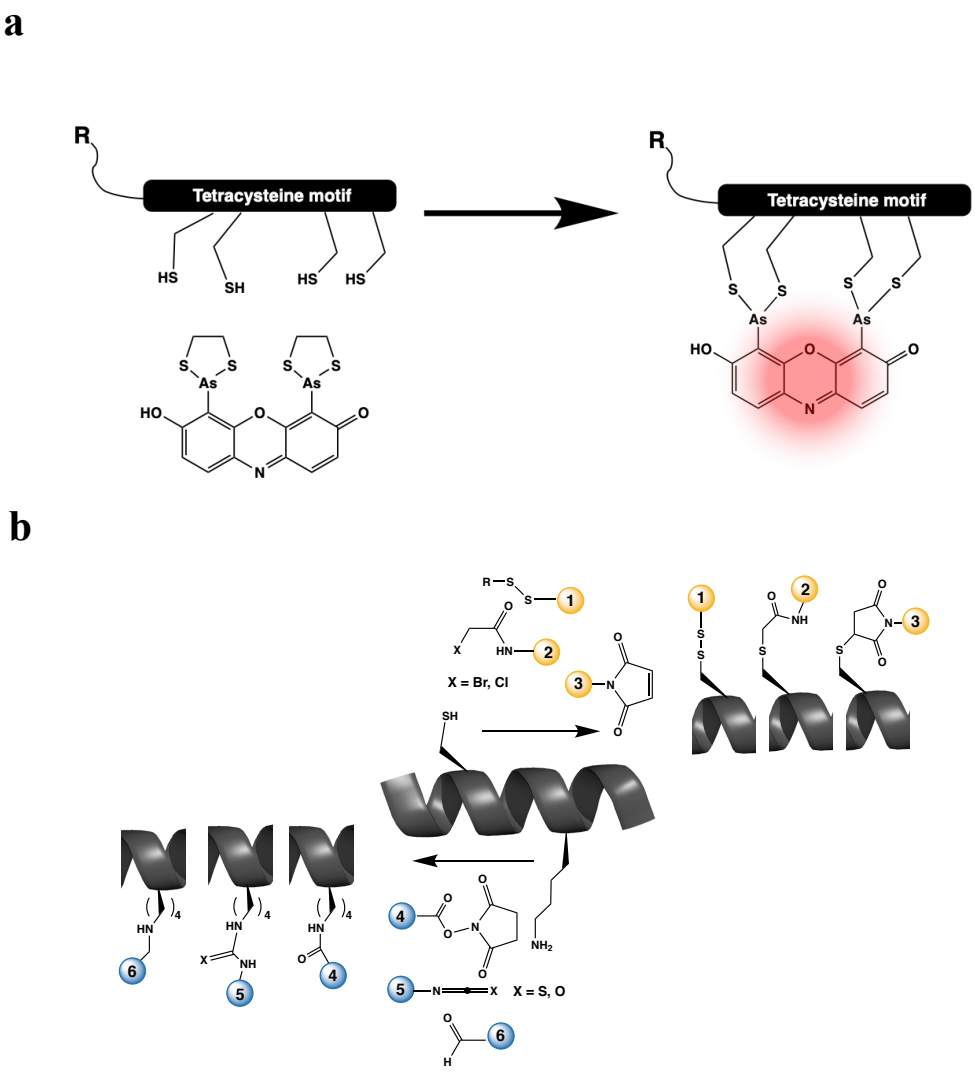


Figure 1.15. Labelling of endogenous amino acids. (a) Reaction of the fluorescent biarsenical compound ReAsH with a tetracysteine tag on a protein. (b) Traditional conjugation strategies at cysteine (yellow) and lysine (blue) (1 = thiol exchange, 2 = alkylation of α-halocarbonyl electrophiles, 3 = maleimide Michael addition, 4 = NHS-ester amidation, 5 = isothiocyanate/isocyanate addition, 6 = reductive amination).

1.3.2. *In vivo* incorporation of non-canonical amino acids

The major limiting factor for the chemical space available for proteins is that the cellular machinery for synthesis is, for the most part, confined to the 20 canonical amino acids. One way that nature has introduced greater functional diversity for proteins is through the myriad of PTMs that can be added after synthesis. The incredible fidelity for the placement of

PTMs is enabled through the evolution of highly specific enzymes capable of distinguishing the appropriate residues in a specific motif, a luxury that chemical biologists do not have for novel modifications. However, artificially expanding the lexicon of amino acids to include ones with novel functional groups can remedy this situation, and numerous approaches to accomplish this have been developed.

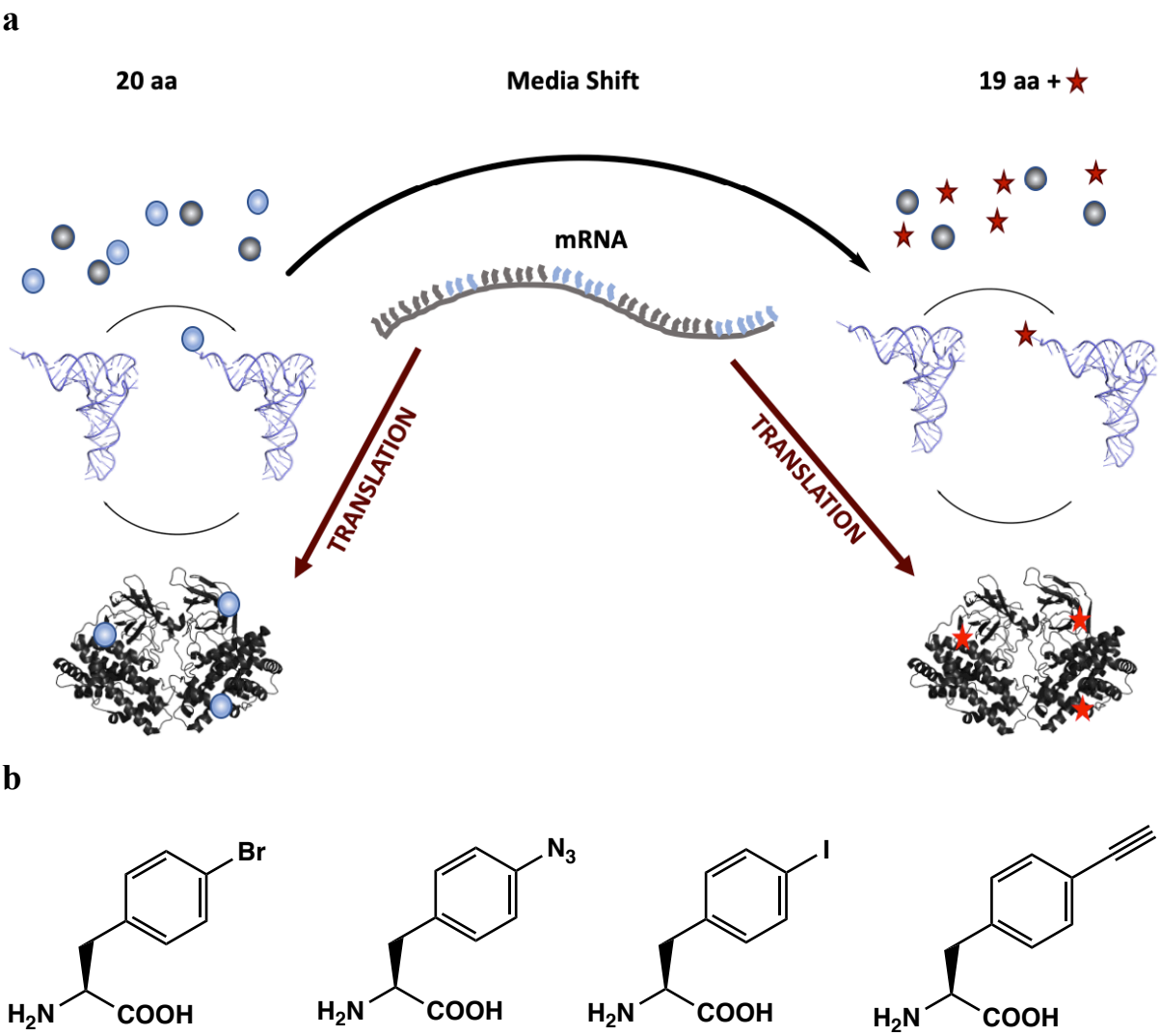


Figure 1.16. Residue specific unnatural amino acid incorporation. (a) Illustration of using an analogue of methionine to and a media shift to incorporate the unnatural amino acid at residue specific sites throughout the protein. (b) Examples of phenylalanine analogues incorporated through the mutation of the endogenous *E. coli* aaRS.

The simplest method for incorporating non-canonical amino acids is through residue-specific incorporation[141]. This approach utilises a chemically similar analogue to a natural amino acid that is able to compete as a substrate for the endogenous aminoacyl-tRNA-

synthetase (aaRS), leading to a partial replacement of the natural amino acid throughout all proteins in the organism (**Figure 1.16a**). An early example of the utility of this approach was back in the 1950s when it was discovered that selenomethionine could be incorporated into bacterial proteins in place of methionine. This facilitated the phasing of many protein crystals thanks to the anomalous scattering that the heavier selenium atom enabled[142]. To achieve closer to quantitative replacement with an unnatural amino acid, auxotrophic strains of bacteria (ones deficient in the biosynthetic machinery necessary to make an amino acid) must be used. Keeping the transgenic protein under a strong repressor, the bacteria can then be grown in rich media with all 20 amino acids initially, and then shifted during expression into media supplemented with the designed analogue replacing the essential amino acid. Without any genetic manipulation, the range of unnatural amino acids that can be introduced in a residue specific manner is limited, as the analogues must be significantly similar to the natural amino acid or else their activation by the endogenous aaRS is too slow or even undetectable. To overcome this, to a certain extent, several different aaRS's have been mutated to be more accommodating to unnatural analogues by modifying either the editing domain (responsible for hydrolysis) or their amino acid binding pocket[143]. In the case of the *E. coli.* phenylalanine aaRS, these types of mutations have enabled the residue-specific incorporation of a variety of uniquely reactive analogues[144] (**Figure 1.16b**). Residue-specific incorporation of reactive tags has enabled studies on the global synthesis, localisation, and degradation of proteins[145,146]. Such indiscriminate incorporation has clear limitations however, and in order to incorporate non-canonical amino acids at specific locations, more advanced methods are needed.

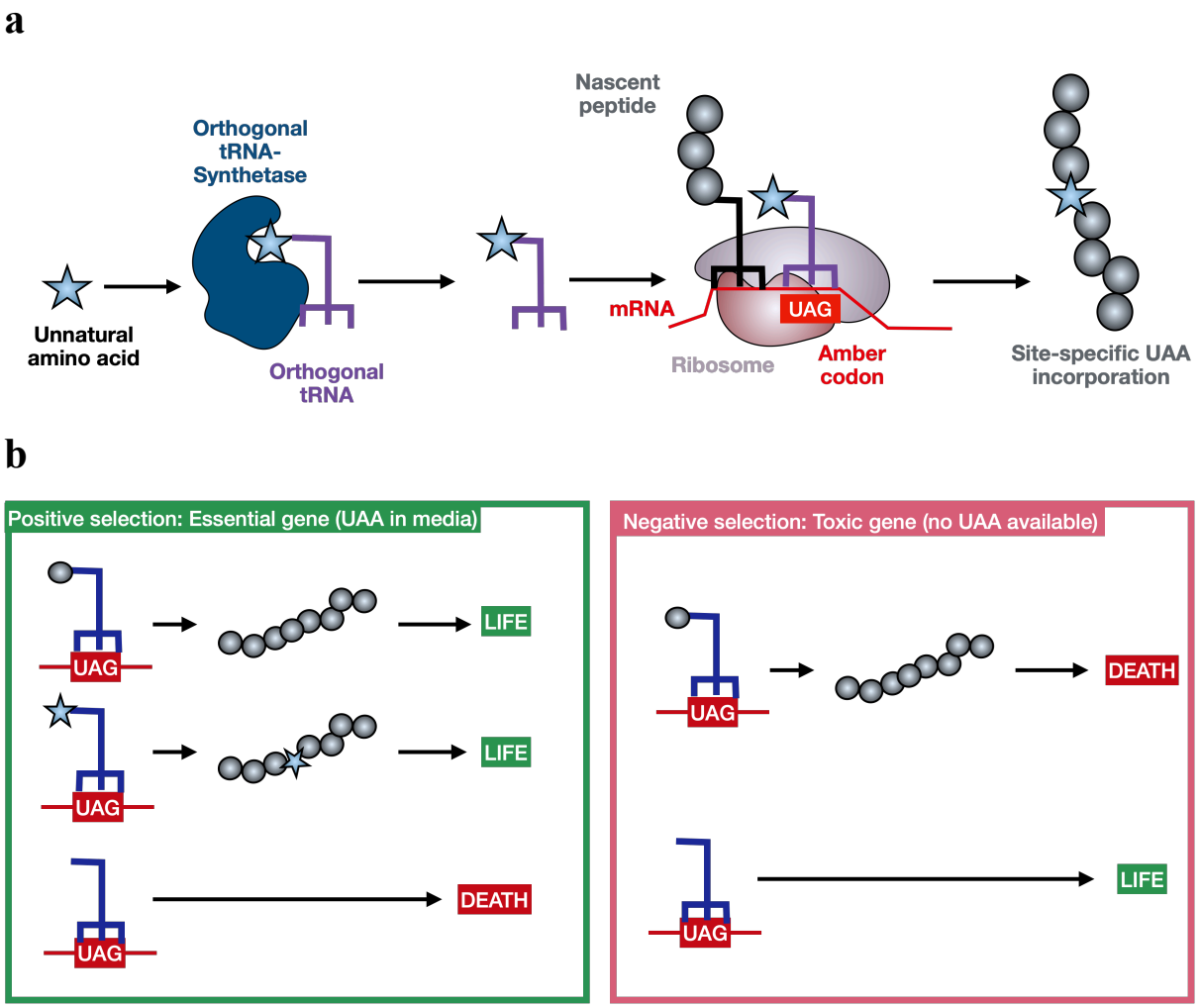


Figure 1.17. Site-specific designer amino acid incorporation. (a) Illustration of the mechanism of amber codon suppression for the expansion of the genetic code. (b) Schematic representation of the positive and negative selection directed evolution steps to engineer an orthogonal aaRS/tRNA pair. The positive step selects for the ability to incorporate an amino acid into an essential protein for life while the negative step reduces promiscuity by removing the designer amino acid from the media and incorporating the amber codon in a toxic protein product.

To site-specifically integrate unnatural designer amino acids, genetic code expansion technology must be used. In this approach, an orthogonal aaRS and tRNA pair are engineered to insert the designed amino acid in response to the amber stop codon in mRNA (UAG) (**Figure 1.17a**)[146]. This approach was pioneered by Peter Schultz and co-workers and has greatly expanded the repertoire of unnatural amino acids. Originally, the approach was made possible by engineering the tyrosine aaRS/tRNA pair from the archaebacterium *M. jannaschii* to recognize the amber codon, integrate unnatural tyrosine analogues, and function orthogonally to the *E. coli* protein synthesis machinery[147]. Systems of this type have now been engineered from multiple organisms to function in prokaryotes all the way to whole multicellular

organisms and have allowed for the incorporation of over 100 disparate unnatural amino acids[141]. The process for developing these systems now has a relatively standard workflow consisting of two directed evolution steps. First, the aaRS/tRNA pair undergo a positive selection step where the unnatural amino acid must be incorporated for the function of a life-giving protein, selecting for any pairs that allow for an amino acid incorporation, and then a negative selection step to reduce the promiscuity of the pair for amino acids other than the desired designer unnatural one (**Figure 1.17b**). The diversity of functional groups that genetic codon expansion technology introduces has enabled a myriad of *in vivo* protein conjugation strategies that have facilitated some significant breakthroughs. A reaction that has proved particularly useful in this context is the [4+2] cycloaddition of 1,2,4,5-tetrazines and various dienophiles, also referred to as the inverse electron demand Diels-Alder reaction (IEDDA) (**Figure 1.18a,b**)[148]. This is due to its biocompatibility, catalyst free reactivity and, in particular, its large second order rate constant. Either the dienophile or the tetrazine can be incorporated into the protein. A realm that this reaction holds particular promise for is therapeutics thanks to the aforementioned qualities, such as for pro-drug activation of antibody-pro-drug conjugates (**Figure 1.18c**)[149]. Compared to conventional ADCs, this approach is markedly safer since the non-specific degradation of the linker simply results in a loss of reactivity rather than the release of the cytotoxic cargo. While genetic code expansion undoubtably has an important role to play for the future of biological research, there is a long road ahead to see the full maturity of this technology. However, great strides are being made to solve issues surrounding the technique, such as the loss of fitness in cells. This is due to read-through of endogenous proteins using the amber stop codon and the challenge of incorporating more than one unnatural amino acid at specific sites. Potential solutions to these issues include the synthesis of a recoded *E. coli* genome with reduced codon usage (61 instead of 64) allowing for the usage of the deleted codons for new amino acids and the creation of a ribosome tRNAs for a quadruplet coding system, increasing the number of codons from 64 to 256[150,151].

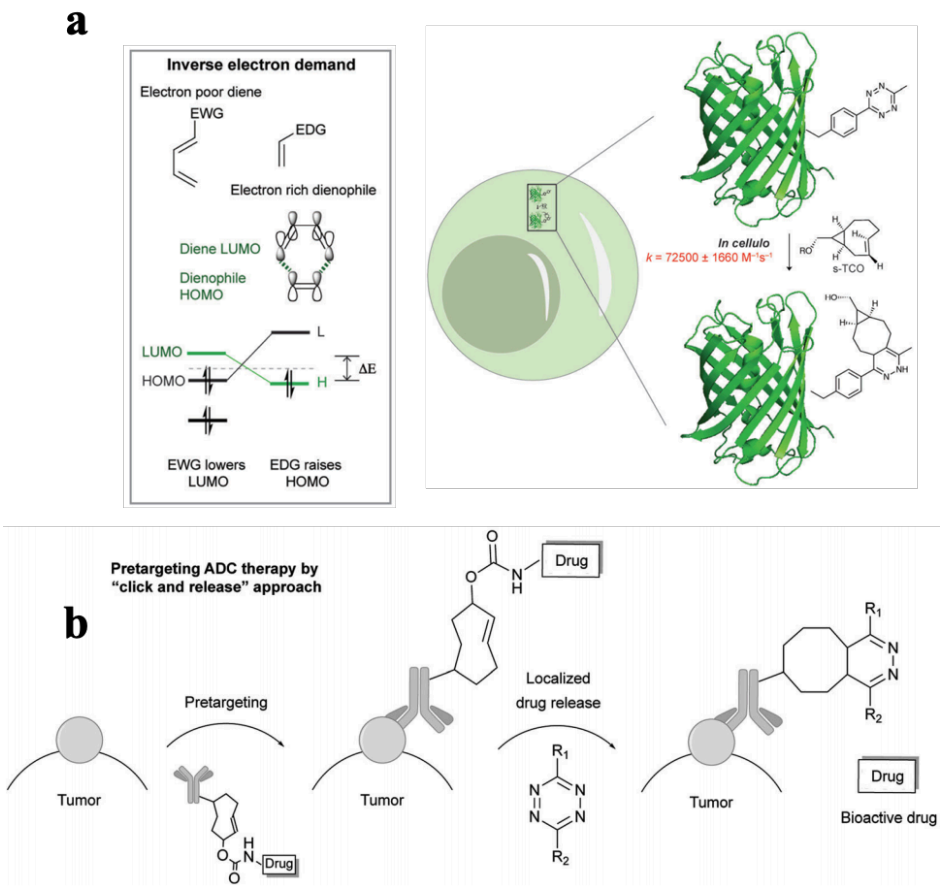


Figure 1.18. IEEDA reaction for *in vivo* bioorthogonal labelling. (a) Frontier molecular orbital diagram of the IEEDA reaction (EDG = electron-donating group, EWG = electron-withdrawing group) and an example of *in vivo* labelling rate of GFP bearing a tetrazine UAA with trans-cyclooctene. (b) Schematic of pro-drug release by an APDC through an IEEDA decoupling reaction. Figures adapted with permission from reference [148].

1.3.3. Post-expression mutagenesis

Site-directed genetic mutagenesis revolutionised molecular biology, especially in the way scientists investigate and modulate protein function. The abovementioned genetic code expansion technology is now further expanding the power of gene-based mutagenesis. However, before these technologies existed, the first ever site-selective point mutation was not achieved through genetic manipulation, but by chemical means on a purified protein. In 1966 both the research groups of Koshland and Bender independently converted the active site serine residue of substilin to cysteine using a two-step chemical synthesis (**Figure 1.19a**)[152,153]. After the 1970s, the rise of recombinant protein technology led many to believe that mutagenesis via chemical means was obsolete. Barring a few notable examples in the context of improving native chemical ligation (NCL) to overcome its reliance on cysteine[154,155], progress in the field

saw a very long hiatus until the mid 2000s. Chemical biologists were driven to reinvestigate chemical mutagenesis by the potential utility of a general approach for post-expression amino acid interconversion. This approach would utilise a precursor residue capable of being transformed into a wide variety of natural and unnatural sidechains, potentially offering a wide array of chemical diversity from a single purification of the precursor bearing protein. Thus, both enhancing the efficiency of production and expanding the array of functional groups available for the protein compared to traditional mutagenesis.

From the description, it is clear that the most important initial step in developing a general chemical mutagenesis method is the selection of a synthetically versatile amino acid. For an amino acid to be considered synthetically versatile it must, by default, have reactivity of a bioorthogonal nature in order to not interfere with other regions of the protein. Genetic code expansion has indeed been used in this context to provide amino acids with this quality. Shultz and co-workers incorporated *p*-borophenylalanine in proteins to act as a traceless affinity tag, as the boronic acid moiety binds to polyhydroxylated resins and can then be converted to either tyrosine or phenylalanine via bioorthogonal means after purification (**Figure 1.19b**)[156]. The access to just two amino acids from a reactive precursor is far from a general approach but highlights a creative application of this technology. To gain true generality, a smaller and more versatile amino acid must be used. As far as general reactivity trends go, none of the canonical amino acids are strong electrophiles. Therefore, the placement of a minimal electrophilic amino acid could be an ideal candidate for a synthetically versatile precursor. To this end, the amino acid dehydroalanine (Dha) has emerged as a useful precursor.

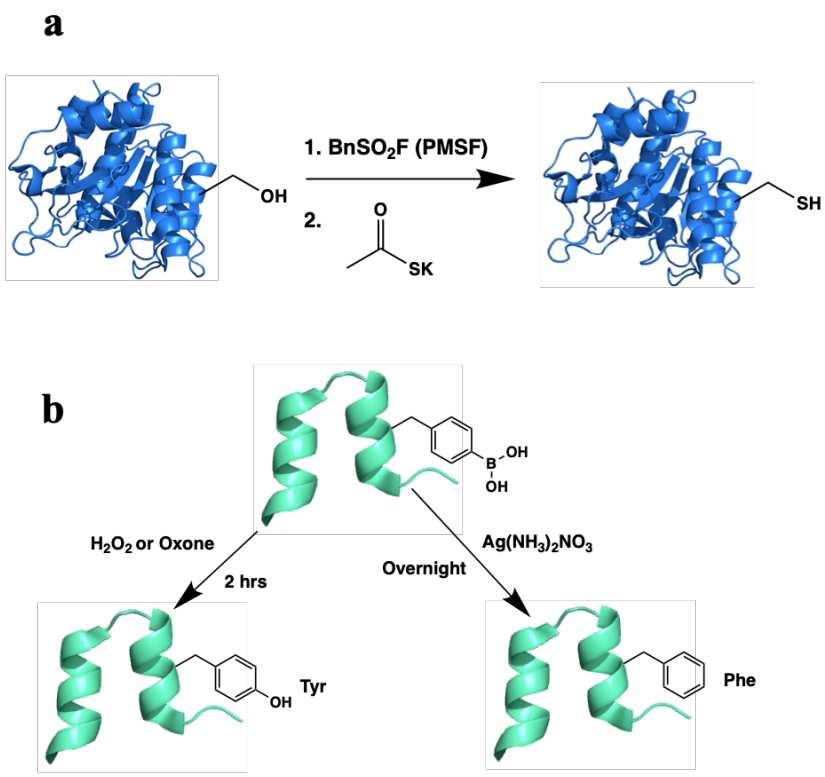


Figure 1.19. Early examples of chemical mutagenesis. (a) Strategy for the first ever site-specific amino acid conversion on substilin, the active site serine was first converted to a strong leaving group that could then be displaced by the thioacetate, which upon hydrolysis yields the free thiol of cysteine. (b) *P*-borophenylalanine as a slightly more diverse chemical precursor to the amino acids phenylalanine and tyrosine.

1.3.4. Dehydroalanine: Installation and modification

The electrophilic nature of Dha has been utilised in several organisms, particularly for the synthesis of complex peptide macrocycles. For example, the bacteria *Lactococcus lactis* uses Dha as an intermediate in the synthesis of the commercially used antibacterial peptide nisin (**Figure 1.20a**)[157]. In nisin's biosynthesis, serine is enzymatically converted to Dha through a phosphorylated intermediate, which is subsequently used as a Michael-acceptor for the creation of a thioether linkage to cyclise the peptide at multiple sites. The α/β unsaturated system of Dha is the unique chemical structure among the canonical amino acids that confers the electrophilicity of the residue and facilitates conjugation through Michael additions and other reactions. While biosynthetic machinery is capable of selectively converting serine residues to Dha, from the perspective of a chemist, this is not a particularly attractive route of installation. Namely, this is because serine is one of the most common amino acids, so selectivity will undoubtably be an issue and the elimination of the hydroxyl group typically requires pH conditions that would be incompatible with many proteins. Again, genetic code

expansion technology has been used to provide an alternative route through the incorporation of phenylselenocysteine, which can be converted to Dha through oxidative elimination[158]. However, an even more easily accessible and appropriate precursor to Dha is cysteine. This is because cysteine is one of the least common amino acids and its site-selective installation does not require an engineered organism. From cysteine, there are four clear possible routes for conversion to Dha: reductive-elimination of cysteine disulphides, base-mediated elimination of disulphides, oxidative elimination of cysteine, and bis-alkylation-elimination of cysteine (**Figure 1.20b**)[159]. Of these strategies, the bis-alkylation-elimination approach has proven to be the preferred route, as it can be accomplished with easily accessible reagents, under mild conditions, without the addition of organic solvents, and on reasonable timescales[159].

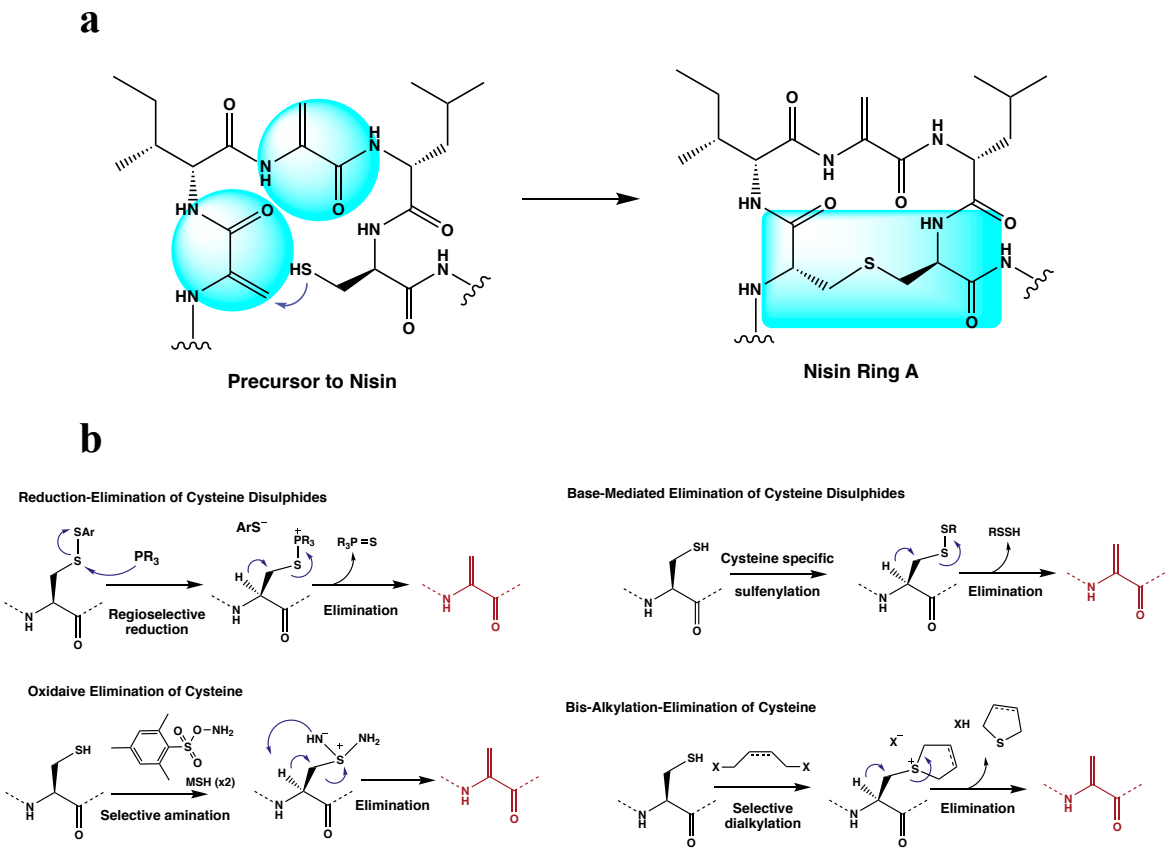


Figure 1.20. Dha in nature and its chemical installation. (a) Synthesis of ring A of Nisin through the Michael addition of a cysteine residue to an enzymatically installed Dha residue. (b) Chemical strategies for the installation of Dha from solvent accessible cysteine residues on peptides and proteins.

Once installed on a protein, the power of Dha for chemical mutagenesis becomes readily apparent from the diverse routes of modification it enables. As mentioned above, the α/β unsaturated system of Dha creates a Michael-acceptor, a unique reactivity among all

canonical amino acids. Using this approach, proteins bearing Dha have been modified through both thia and aza-Michael additions with a plethora of sidechains[159-162]. Due to the minimal linking group resulting from the addition (a carbon and a single heteroatom), many complex PTM mimetics have been able to be installed on proteins using Dha to enable the accurate study of their function, despite the fact that modifications at Dha likely result in an epimeric mixture of the installed sidechain (**Figure 1.21a**). For example, Davis and co-workers used Dha mediated synthetic phosphorylation of the disease relevant kinase p38α to accurately study therapeutic inhibitors[163]. The activity of p38α is regulated by phosphorylation at Thr180, and it is difficult to activate p38α non-enzymatically as typical phospho-mimetic mutations like Asp/Glu do not activate p38α to its endogenous substrate ATF2[164]. By creating a cysteine mutant at Thr180 and then converting to Dha, the researchers were then able to create a phosphocysteine (pCys) residue through a Michael addition with sodium thiophosphate, a very close analogue to pSer. The resulting p38α(Thr180pCys) mutant was capable of phosphorylating ATF2 *in vitro* to levels in line with studies using biologically derived mixed samples (> 10% non-phosphorylated and ~30% diphosphorylated) (**Figure 1.21b**)[163]. Closer analysis of the time course by LC-MS revealed that the synthetically mono-phosphorylated form also favoured mono-phosphorylation of ATF2, while the mixed samples containing di-phosphorylated p38α favoured di-phosphorylation of ATF2, thereby shedding light on a previously unknown mechanism. Creation of the pure active form through synthetic means also enabled the unequivocal assignment of modes of inhibition (i.e. Type I/II) to a panel of inhibitors and quantify their cooperativity. Michael-additions at Dha also hold great promise in the realm of biotechnology. Bernardes and co-workers successfully used Dha to create an ADC with Thiomab (anti-Her2 antibody) and Crizotinib (FDA approved kinase inhibitor) by ligating the drug directly to the antibody through an aza-Michael addition of the piperidine moiety within the drug itself[161]. This created product with a homogenous drug-to-antibody ratio of 2 and a plasma stable tertiary amine as the linker and furthermore improved the cell-killing activity of the drug ten-fold.

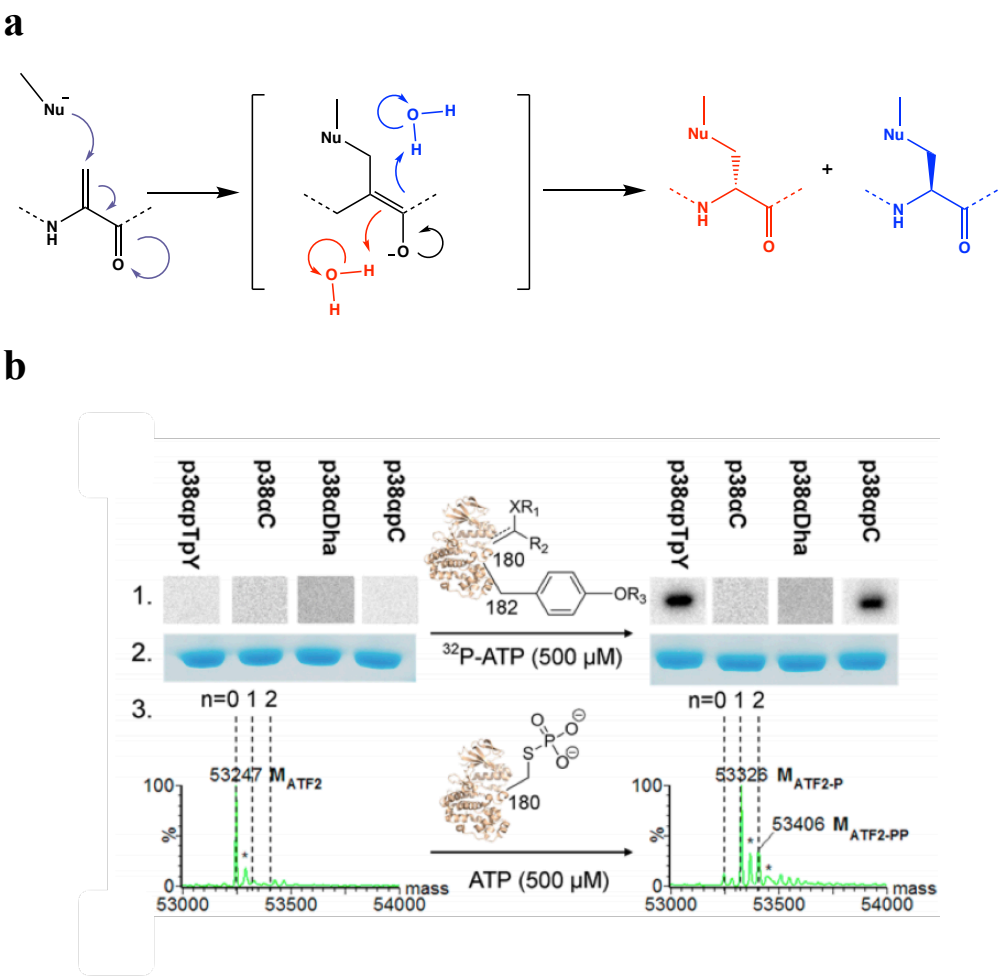


Figure 1.21. Michael additions at Dha and mimicking a disease relevant PTM. (a) Epimerization upon the nucleophilic attack of Dha by the proton transfer from water to the enolate intermediate from above and below. (b) 1- Electrophoretic radioassay of ATF2 phosphorylation by various forms of P38α, showing that the synthetic phosphorylation (P38αpC) is a valid mimic. 2- Coomassie stain verifying the total ATF2 levels. 3- LC-MS quantification of the degree of ATF2 phosphorylation, illustrating that the mono-phosphorylated P38α favours mono-phosphorylation of its target. Figure (b) adapted with permission from reference [163].

While the literature supports that Michael additions at Dha with thiols and amines has its clear utility in studying PTMs, chemists have developed more advanced reactions to install even more accurate mimics. It is clear that this necessitates a reaction that can create carbon-carbon sp^3 bonds at the β carbon of Dha. To accomplish this, Davis and co-workers have developed a radical mediated ligation of halogen bearing precursors to Dha (**Figure 1.22**)[165]. This approach has enabled the installation of a wide variety of exact PTMs found in nature (except for their epimeric mixture). While this reaction represents an exciting step forward towards a general method for chemical mutagenesis, its complicated mechanism and need for

anaerobic conditions means considerably more optimisation is needed for each reaction compared to the aforementioned Michael additions.

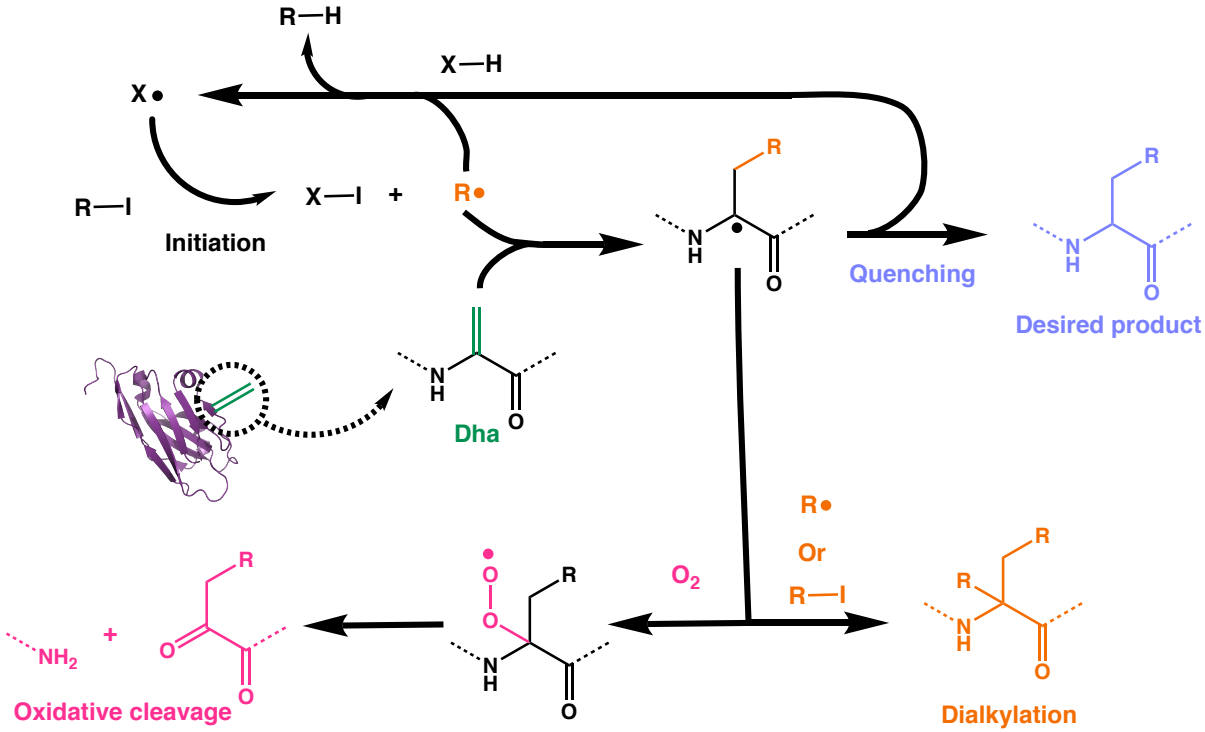


Figure 1.22. C(sp^3)-C(sp^3) bond formation at Dha. Schematic of the possible reaction routes for a free radical mediated C(sp^3)-C(sp^3) bond formation at Dha with halogen bearing precursors.

1.4 Thesis aims

This research was designed to explore the application chemical mutagenesis to both engineer protein activity and study natural PTMs in the context of Alzheimer's disease. Chapter 2 will discuss efforts to use chemical mutagenesis to enhance the anti-aggregation activity of a nascently functional *in silico* designed single-domain antibody for $A\beta_{42}$. This consists of the investigation of the properties of Dha at relevant sites of the protein with the overall aim to provide a new route to mature antibody activity in a rational manner similar to structure activity relationship (SAR) studies conducted routinely on small molecules. With the rapidly expanding repertoire of reactions for site-selective protein modification, we believe this field is ripe for the expansion of its applications to more ambitious purposes such as the one above.

Chapter 3 will focus on the exploration of the use of chemical mutagenesis to provide easy access to diverse tau PTMs. The main goal is to establish protocols for the efficient and homogenous installation of functional and disease relevant PTMs and validate their physiological relevance. Tau PTMs have long been mimicked crudely by genetic approaches,

and with the accuracy of PTM mimics installed via chemical mutagenesis verified for many proteins, we believe that Tau research is poised to benefit greatly from this accurate and simple approach. The overall findings and implications will be discussed in Chapter 4 along with recommended future work.

Chapter 2: Systematic anti-aggregation enhancement of an Aß$_{42}$ single-domain antibody through chemical mutagenesis

2.1 Introduction

2.1.1. Chemical activity maturation of proteins

Small-molecule (SM) drug discovery has very different but powerful methods for maturing biological activity of starting candidates when compared to antibodies. With the synthetic power of modern medicinal chemistry, initial SM hits can have various aspects of their biological activity matured atom by atom through structure activity relationship (SAR) studies. SAR studies can be applied to increase the potency of SMs, improve their solubility, enhance bioavailability, and reduce off-target effects[166]. Replicating the accuracy and efficiency of SM SAR studies with proteins would be highly desirable, as current methods of directed evolution and rational mutagenesis typically lead to trade-offs of potency and stability (or vice-versa), especially when the target is aggregation prone[116]. To truly replicate the power of SAR studies for proteins, the limited chemical diversity of the canonical proteinogenic amino acids must be overcome. Unsurprisingly, genetic code expansion technology has been explored to address this. Schultz and co-workers carried out 4 different phage display campaigns against the same target (HIV protein gp120) each with a different 21st amino acid to expand the chemical space[167]. This work did indeed find that the candidate with the highest affinity for gp120 contained a sulfotyrosine residue in the V_H CDR3 loop of the scFv. However, this approach is limited by both the fact that the genetic code expansion can only increase the chemical lexicon one amino acid at a time and that it inevitably relied on a randomised evolutionary process. To efficiently perform a SAR study, ideally there would be a post-expression system for the facile and rapid expansion of the chemical space at key residues.

In this spirit, Berry and co-workers have utilised Dha as a synthetically versatile precursor to enhance enzyme activity and specificity[168]. Within the active site of the promiscuous aldolase *N*-acetylneuraminic lyase (NAL), the researchers screened 13 non-canonical side chains at 12 different positions within the active site accessed through thia-Michael additions at Dha placed at the various sites. This work identified that one chemical

mutant (F190-2,3-dihydroxypropyl cysteine) enhanced the k_{cat} of NAL for a specific substrate pair 10-fold over any natural amino acid. This application of chemical mutagenesis for protein engineering represents an exciting development in the field and could enable catalytic functions not found in nature. For reasons discussed above, antibodies represent a very attractive next target for the application of chemical activity maturation. To this end, this chapter will explore the various chemistries available for chemical mutagenesis for a rationally designed single-domain antibody against the N-terminus of $A\beta_{42}$ with nascent anti-aggregation activity with the aim to find hits that will improve this activity.

2.1.2. Chapter aims

While affinity maturation of antibodies can be accomplished through directed evolution methods, enhancing more complex biological functions is difficult to accurately accomplish using these techniques. These functions include creating agonists and antagonists for signalling pathways through membrane protein engagement, inhibiting proteins crucial to the pathogenesis of viruses and bacteria through specific epitope engagement, as well as the inhibition of protein aggregation. Simply binding an immobilised target in these contexts does not necessarily impart the desired function. Therefore, after traditional display methods each candidate needs to be screened in a subsequent functional assay, which is time-consuming, tedious and costly. The overarching aim of this chapter is to explore if chemical mutagenesis can be an efficient avenue for the accurate functional maturation of an antibody candidate through a SAR like approach (**Figure 2.2**). The starting single-domain antibody contains a rationally designed peptide through the cascade method described in section **1.2.4** grafted into a human single domain V_H scaffold containing thermally stabilising mutations to encourage fidelity through the reaction sequences[116] (termed DesAB-$A\beta_{3-9}$). This single-domain antibody was chosen as our initial construct because of its starting inhibitory activity (**Figure 2.18a,b**). DesAB-$A\beta_{3-9}$ inhibits the primary and secondary nucleation pathways to a nearly equal extent. Therefore, we were curious to determine if an activity maturation campaign would enhance the inhibition of both pathways simultaneously or if one would be preferred over the other. A previous graduate student had also attempted conventional phage-display techniques on this antibody, which only yielded candidates with grossly compromised thermal stability. Within this chapter, firstly the chemistry to facilitate Dha mediated chemical mutagenesis will be explored along the CDR3 loop screening for amenable strategies, positions and conditions.

Then a first panel of diverse chemical mutants will have their desired activity investigated in a thioflavin-T (ThT) $A\beta_{42}$ anti-aggregation screen. Following the principles of SAR studies, an initial chemical mutant hit will then be further derivatised and characterised. Some of the chemistry of this project was done with assistance from Robertinah Rakoto, an Erasmus student from University of Strasbourg, from February 2018 to April 2018.

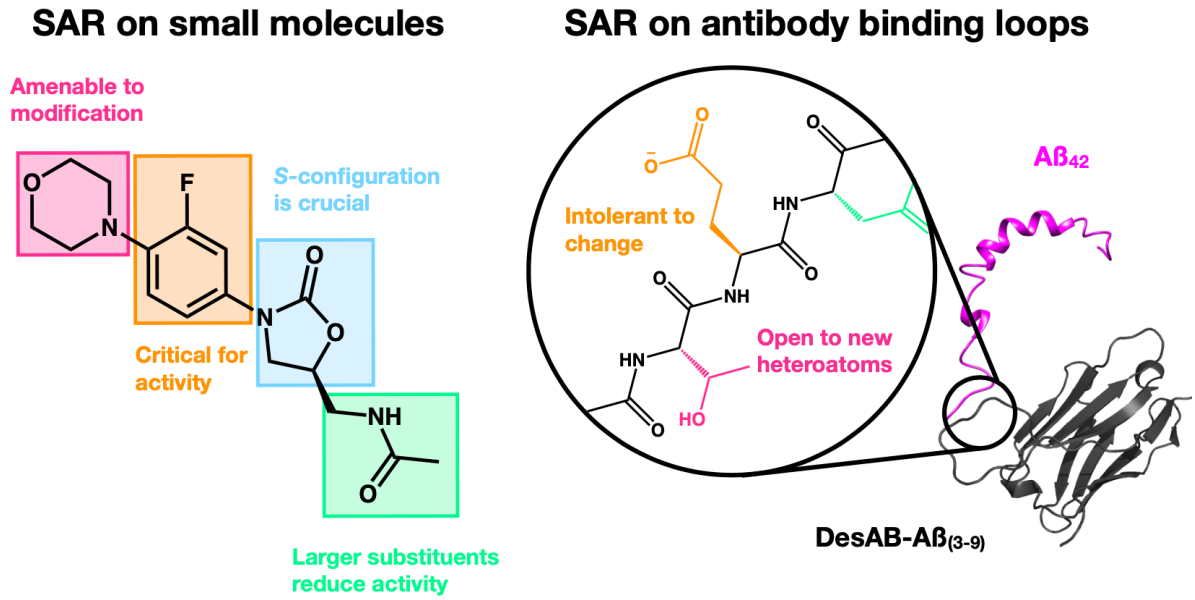


Figure 2.1 SM SAR and SAR on an antibody binding loop. Illustration of the consequences of chemical modifications revealed through a SAR on an oxazolidinone antibiotic compared to the planned SAR study on an antibody binding loop.

2.2 Results

2.2.1. Cloning and purification of DesAB-$A\beta_{3-9}$ and the CDR3 loop cysteine mutants

In order to introduce Dha into a protein, a precursor amino acid must be decided upon that can be chemically converted to Dha. As discussed in section **1.3.4**, cysteine represents a very attractive candidate for this purpose as it can be incorporated through conventional site-directed mutagenesis, and in the case of DesAB-$A\beta_{3-9}$, there are only two native cysteines that are engaged in a disulphide bond within the interior of the protein. The starting thermostable DesAB-$A\beta_{3-9}$ DNA sequence was in a pRSET-B vector with a cleavable hexa-histidine tag to facilitate purification. To introduce cysteine mutants into the CDR3 loop site-directed mutagenesis was carried out at described in ref [169] (see **Chapter 5 Materials and Methods**). The method utilises efficient partially overlapping mutagenic primers that were designed to promote primer-template annealing, eliminate primer dimerization and permit newly

synthesised DNA to be used as a template. This approach facilitated the efficient creation of the cysteine mutants along the 7-residue CDR3 loop (H136C, E137C, T138C, L139C, T140C, L141C, and R142C).

With the cysteine mutant plasmids in hand the panel of proteins were then transformed into BL21 (DE3) Gold cells and expressed and purified according to a standard protocol (see **Chapter 5 Materials and Methods**). Briefly, bacterial cultures were grown to an $OD_{600\ nm}$ of 0.6 and then expression was induced by the addition of 1 mM IPTG. The cultures were left to express overnight at 28 ºC and then lysed via sonication. The antibodies were initially purified from the lysate by immobilised metal affinity chromatography using Ni-NTA beads and then further purified by size exclusion chromatography (SEC) (**Figure 2.3a-c**). This protocol produced each mutant in excellent yield and purity (**Figure 2.2c**).

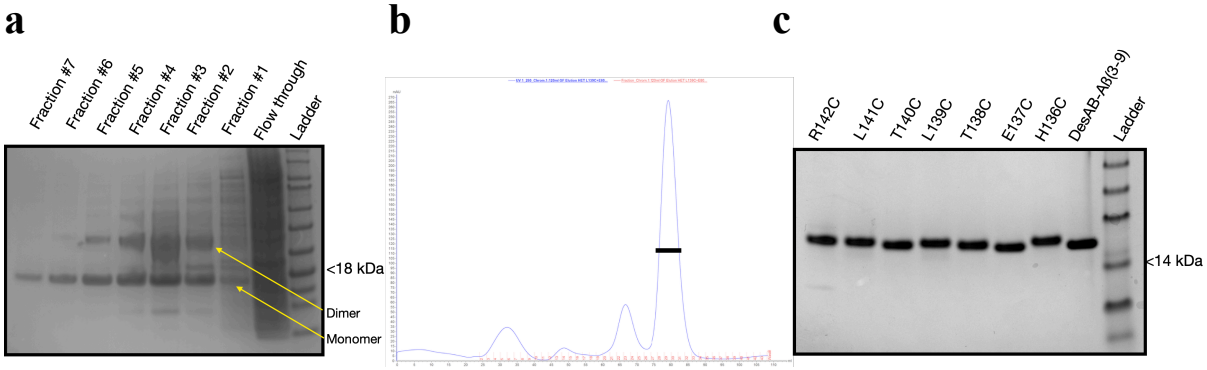


Figure 2.2. Purification of DesAB-Aβ$_{3-9}$ mutants. (a) Representative SDS-PAGE gel for various elution fractions from the Ni-NTA resin for DesAB-Aβ$_{3-9}$ (E137C), using no reducing agent shows the formation of a population of dimerised protein. (b) Chromatogram of a representative SEC run of DesAB-Aβ$_{3-9}$ (E137C), fractions under the black bar were collected and pooled. (c) SDS-PAGE under reducing conditions of a DesAB-Aβ$_{3-9}$ and all the Cys mutants used in this study.

2.2.2. Dha installation along the CDR3 loop

With the relevant cysteine mutants in hand the next phase was to explore the ability to install Dha at the various sites. As discussed in section **1.3.4**, there are many potential methods to introduce Dha from cysteine. Chalker and co-workers had extensively explored the compatibility and efficiency of the established methods for installation from cysteine and found that a bis-alkylation elimination strategy using 2,5-dibromohexanediamide was the most widely useful approach[159] (**Figure 2.4**). Crucial to the current study, the researchers had used this method on a disulphide containing camelid nanobody to perform a site-selective

glycosylation reaction and found the reaction to effectively alter the sole free cysteine without the need to for an organic co-solvent. Inspired by their success, we attempted similar conditions to install Dha mutants along the CDR3 loop of the cysteine mutants. To briefly summarize their approach, proteins were placed in a mildly basic buffer (usually sodium phosphate [NaP_i], pH 8) as it is crucial to have a substantial population of thiolate anions to drive the initiation of the reaction. The pKa of Cys typically ranges from 8-9, therefore at pH 8, 50-10% will be in the thiolate state respectively, enough to drive the reaction and typically tolerated by proteins. The rate limiting step in the formation of Dha using 2,5-dibromohexanediamide has been observed to be the final elimination of the proton from the α-carbon, driving the expulsion of the cyclised sulphur moiety[159]. To encourage the completion of this final part of the reaction, it has been found to be crucial to run the reaction at 37 °C[159].

To break up any dimers that may have formed after the SEC step, 100 µL aliquots of 75 µM proteins in 20 mM NaP_i (pH 8) buffer were incubated with 100 equivalents of DTT for 30 mins at room temperature (RT) with gentle shaking. DTT was then removed by a 7 kDa MWCO zeba desalting column using the manufacturer's instructions and added directly to 500 molar equivalents of powdered DBHDA and incubated at 37 °C shaking at 500 rpm for 5 hrs. Aliquots of each reaction were taken after the incubation and examined by LC-MS, confirming the homogenous formation of Dha at all positions (**Figure 2.4**).

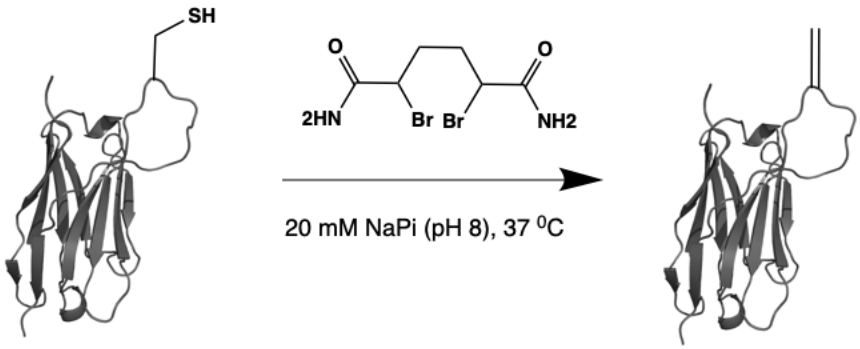
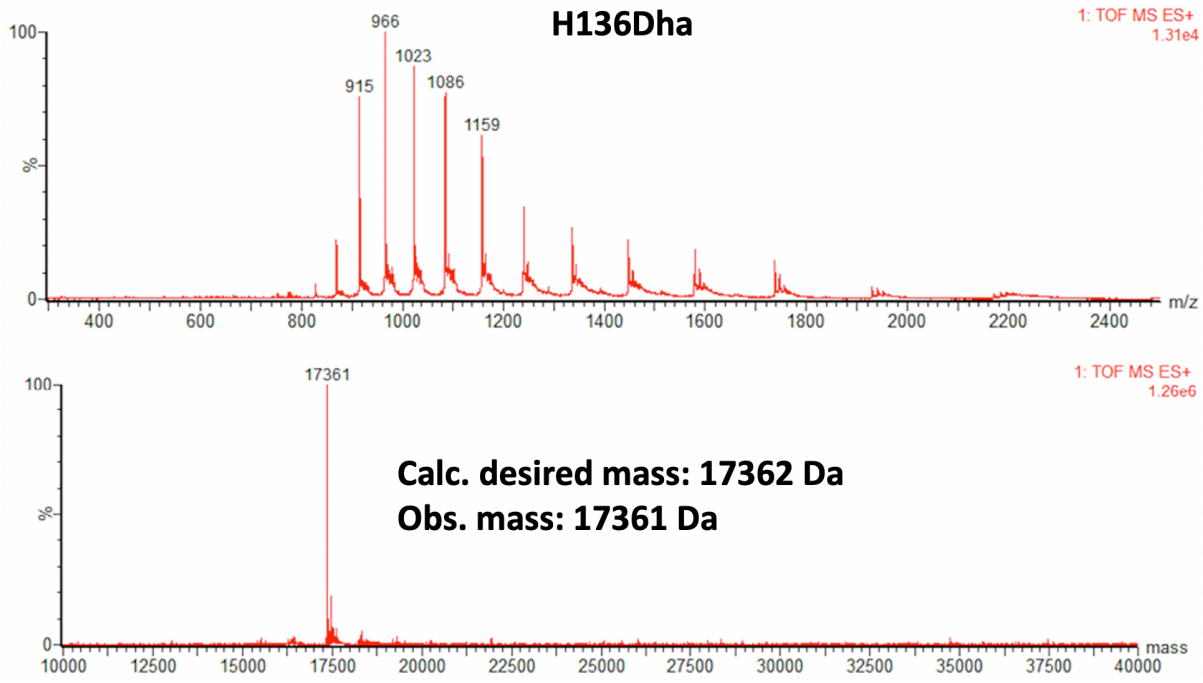
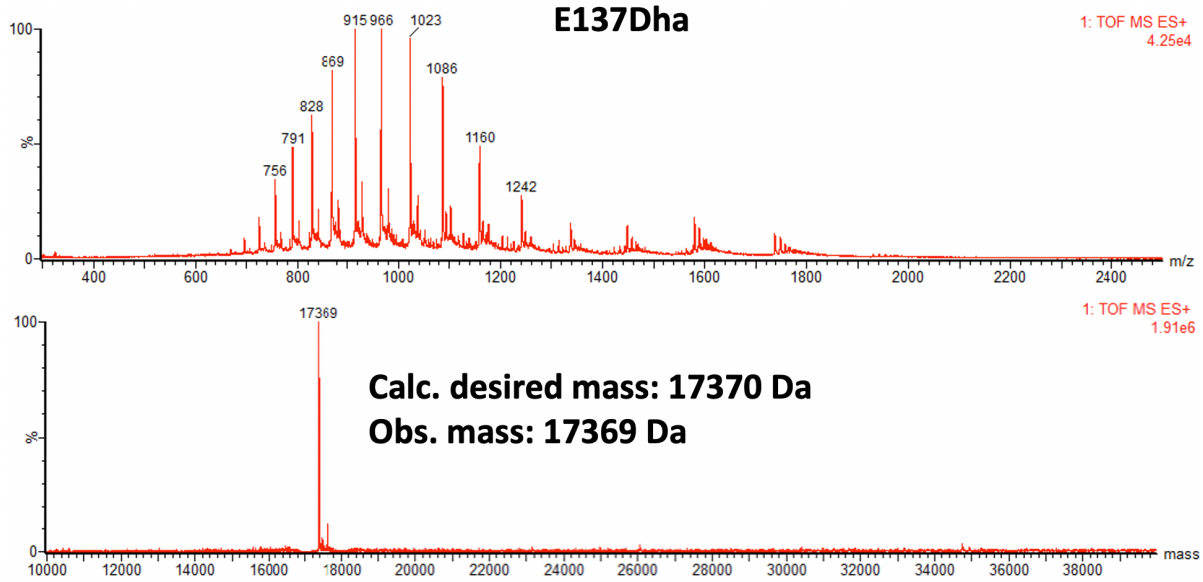


Figure 2.3. Dha conversion along the CDR3-loop of DesAb-Aß(3-9). To a 200-500 µL aliquot of reduced protein (~100 µM), 500 molar equivalents of solid DBHDA was added and the solution was vortexed for 30 seconds and then shaken at 500 rpm at 37 °C. The reaction was monitored by LC-MS with time points taken at 3 or 5 h (see panels a-g below). Time points were taken by aliquoting 2 µL of the reaction mixture and diluting it with 18 µL of the reaction buffer (20 mM NaPi (pH 8)), after which 5 µL of the diluted samples were injected into the LC-MS. Full conversion of the starting cysteine mutant to dehydroalanine was observed at all sites after 3-5 h. Excess DBHDA was removed by centrifuging the samples (10 min, 10,000g, 4 °C).

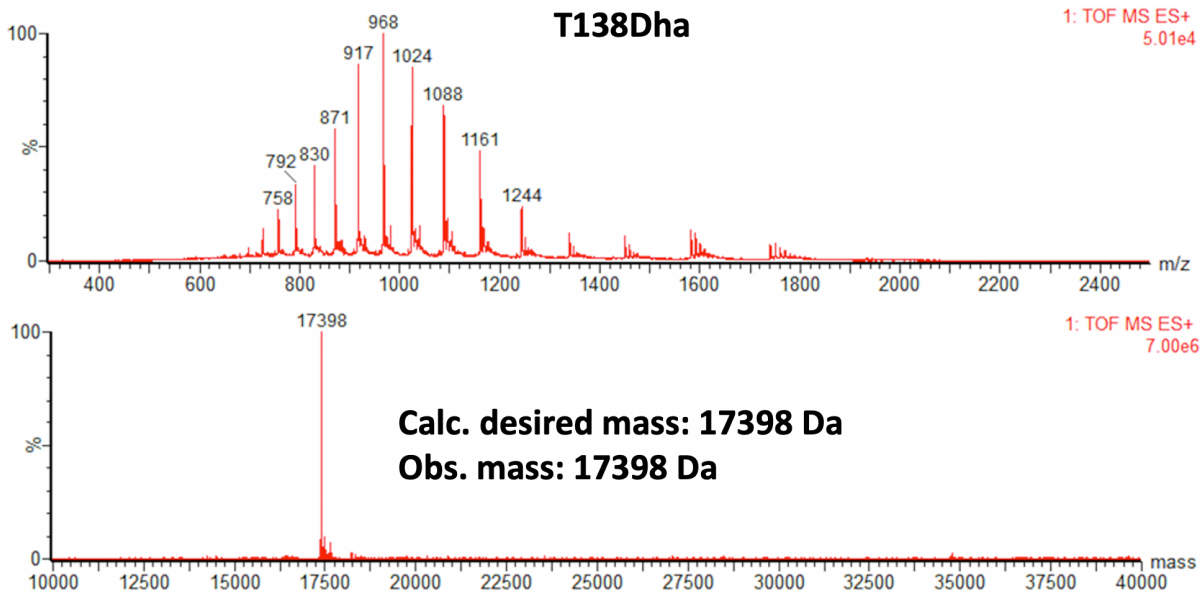
a


H136Dha
1: TOF MS ES+
1.31e4
966
1023
915
1086
1159
100
%
0
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
17361
1: TOF MS ES+
1.26e6
Calc. desired mass: 17362 Da
Obs. mass: 17361 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass

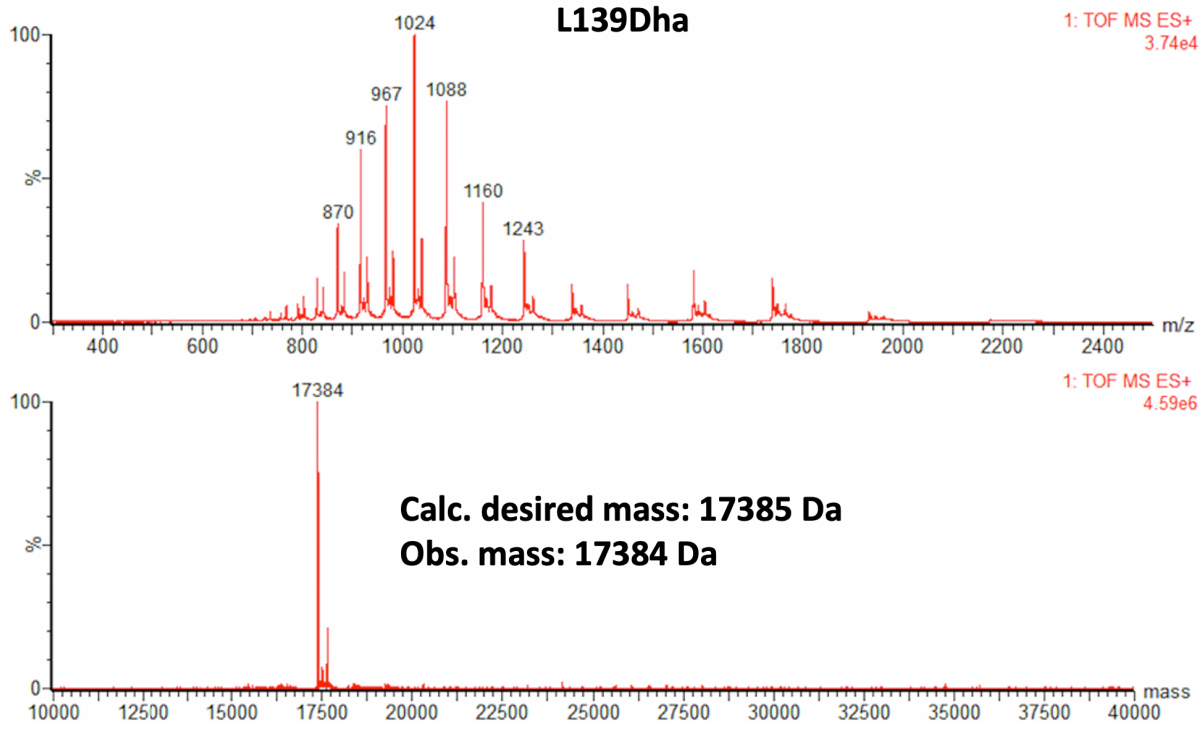
b


E137Dha
1: TOF MS ES+
4.25e4
915 966
1023
869
1086
828
791
1160
756
1242
100
%
0
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
17369
1: TOF MS ES+
1.91e6
Calc. desired mass: 17370 Da
Obs. mass: 17369 Da
10000
12000
14000
16000
18000
20000
22000
24000
26000
28000
30000
32000
34000
36000
38000
mass

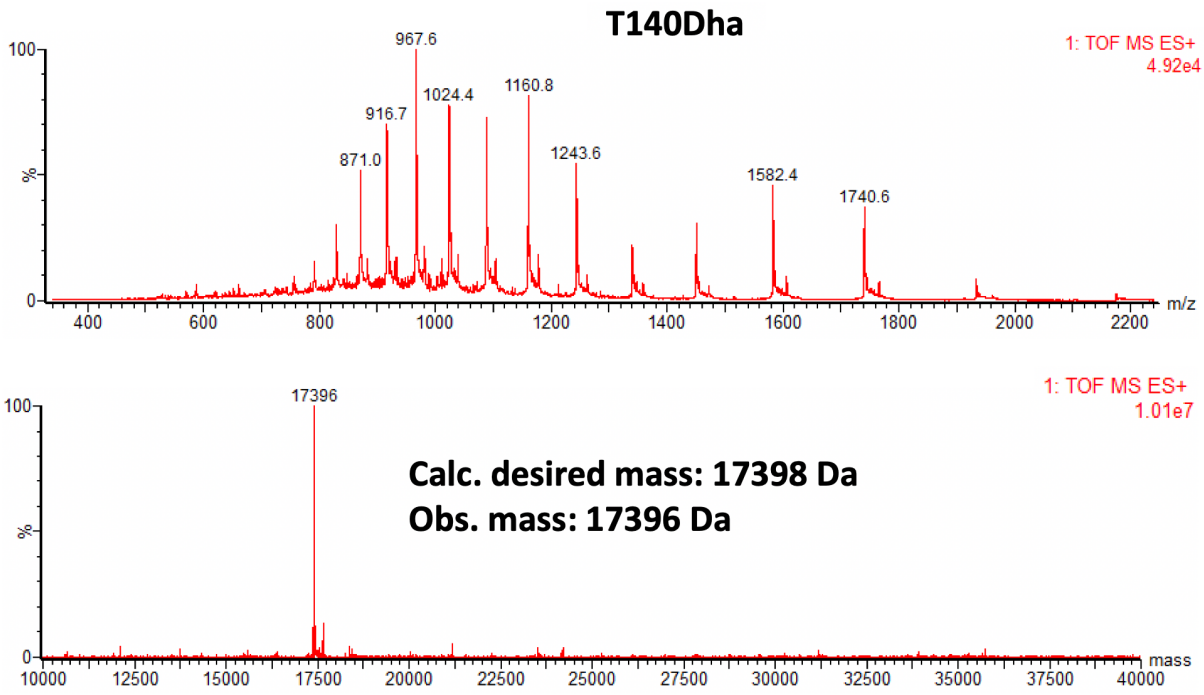
c


T138Dha
1: TOF MS ES+
5.01e4
968
917
1024
1088
871
1161
792
830
758
1244
m/z
1: TOF MS ES+
7.00e6
17398
Calc. desired mass: 17398 Da
Obs. mass: 17398 Da
mass

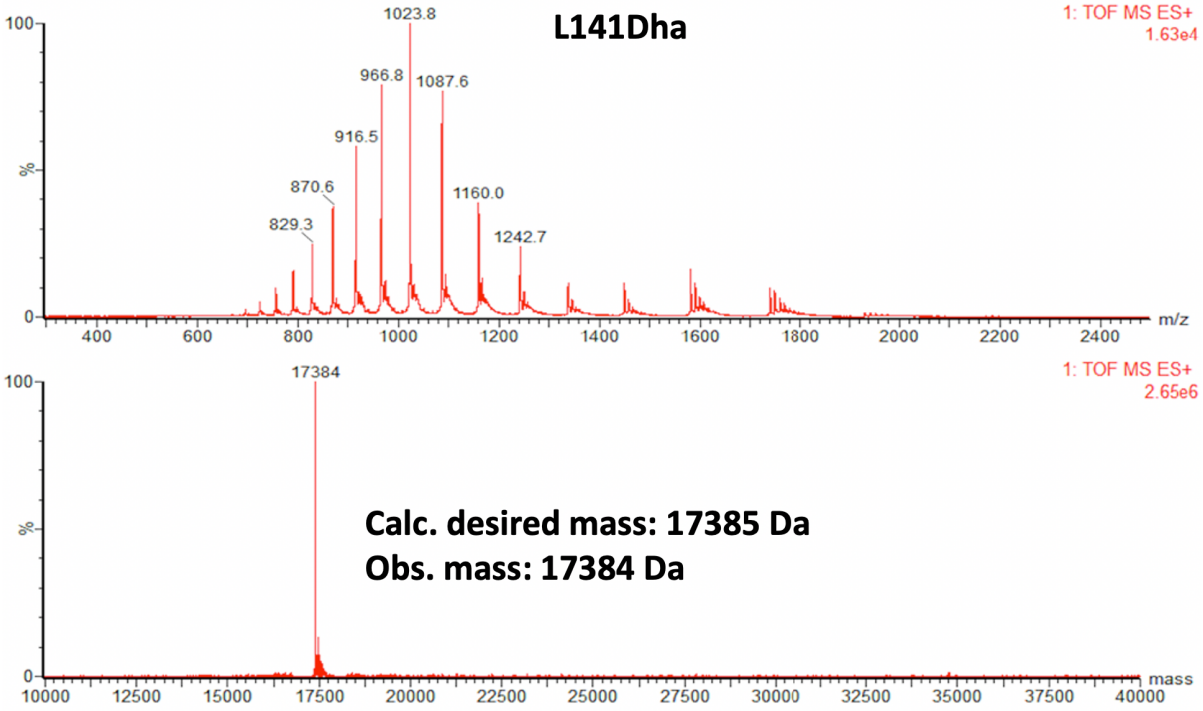
d


L139Dha
1: TOF MS ES+
3.74e4
1024
967
1088
916
1160
870
1243
m/z
1: TOF MS ES+
4.59e6
17384
Calc. desired mass: 17385 Da
Obs. mass: 17384 Da
mass

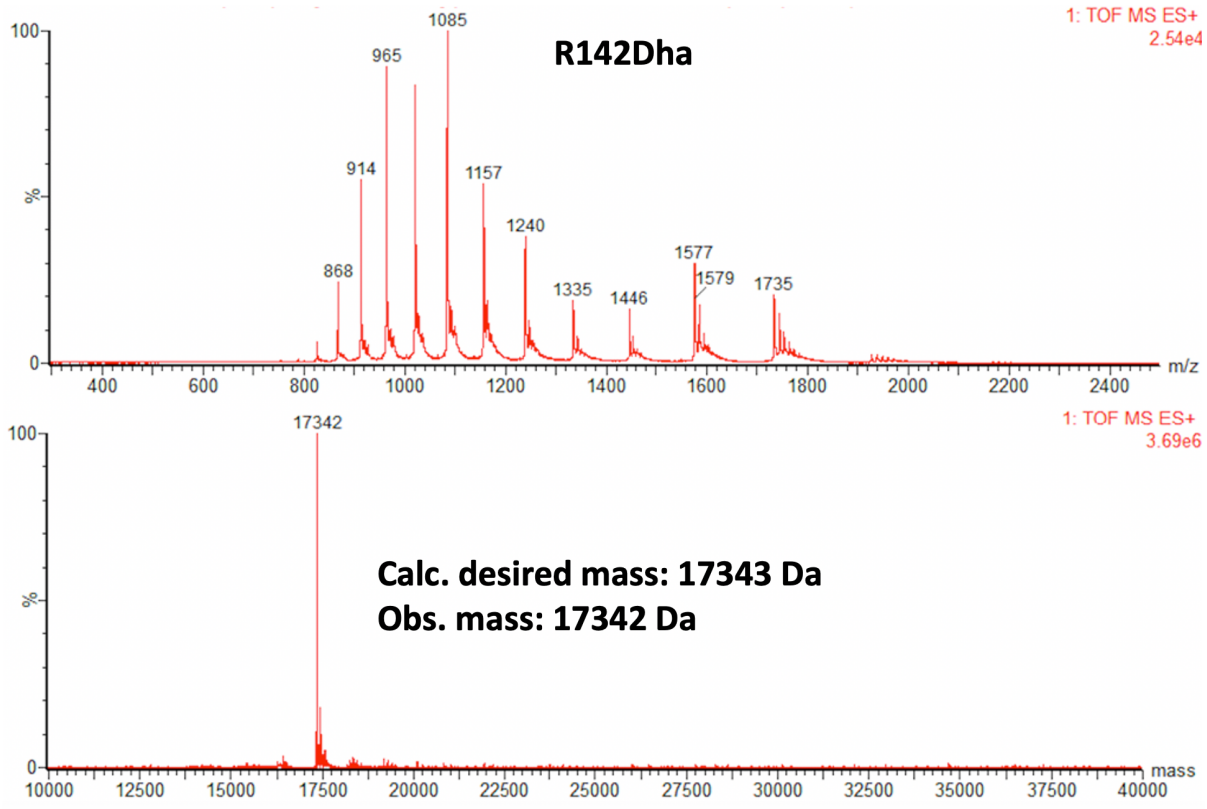
e


T140Dha
1: TOF MS ES+
4.92e4
967.6
1024.4
916.7
1160.8
871.0
1243.6
1582.4
1740.6
m/z
1: TOF MS ES+
1.01e7
17396
Calc. desired mass: 17398 Da
Obs. mass: 17396 Da
mass

f


L141Dha
1: TOF MS ES+
1.63e4
1023.8
966.8
1087.6
916.5
870.6
829.3
1160.0
1242.7
m/z
1: TOF MS ES+
2.65e6
17384
Calc. desired mass: 17385 Da
Obs. mass: 17384 Da
mass

g



2.2.3. Mapping Dha reactivity with β-mercaptoethanol

With Dha mutants successfully installed at each position, the reactivity of Dha at the various positions was evaluated to establish the most practical ones to move forward with side-chain exploration. Most studies using Dha verify its incorporation by observing its reactivity with β-mercaptoethanol via LC-MS. Inspired by this standard in the field, we tracked the addition of β-mercaptoethanol at each position by LC-MS in order to accurately track the time to completion as a measure of general amenability to modification (**Figure 2.3**). Each Dha mutant at 50 μM in 20 mM NaP_i (pH 8) was reacted with 1000 molar equivalents of β-mercaptoethanol in a total final volume of 100 μL at 37 °C shaking at 500 rpm. Samples were removed and analysed at either 15 min or 1 hr timepoints depending on the speed of the reaction. This assessment revealed that while Dha incorporation at the various positions was rather uniform, the subsequent reactivity of Dha is heterogeneous and highly dependent on the local chemical environment. Some positions reached completion (>95% conversion) in <15 min (E137Dha and L139Dha) while others took as long as 6 hours (R142Dha). While it was not the intention of this experiment to identify the sequence contexts that make Dha reactive

or not, it is possible to make some reasonable speculations from this data. The least reactive position (R142Dha) is directly adjacent to three successive Glu residues on its C-terminal side. It appears that the proximity to negatively charged residues (Glu) hinders the addition of β-mercaptoethanol, possibly due to unfavourable electrostatic interactions for the attacking thiolate anion. Whatever the reason, the three most promising positions to take forward for further side chain exploration appeared to be E137Dha, T138Dha, and L139Dha for their timely modification by β-mercaptoethanol and because they represent three distinct amino acid types (charged, polar, and nonpolar). To completely verify that these were indeed the sites of modification and not the cysteines in the disulphide position, these three positions had the β-mercaptoethanol addition mapped by tandem mass spectrometry (LC-MS/MS) that was gratefully performed by Dr. Mike Deery of the Department of Biochemistry at Cambridge (**Figure 2.4**). This investigation indeed verified that the β-mercaptoethanol addition occurred at the intended sites.

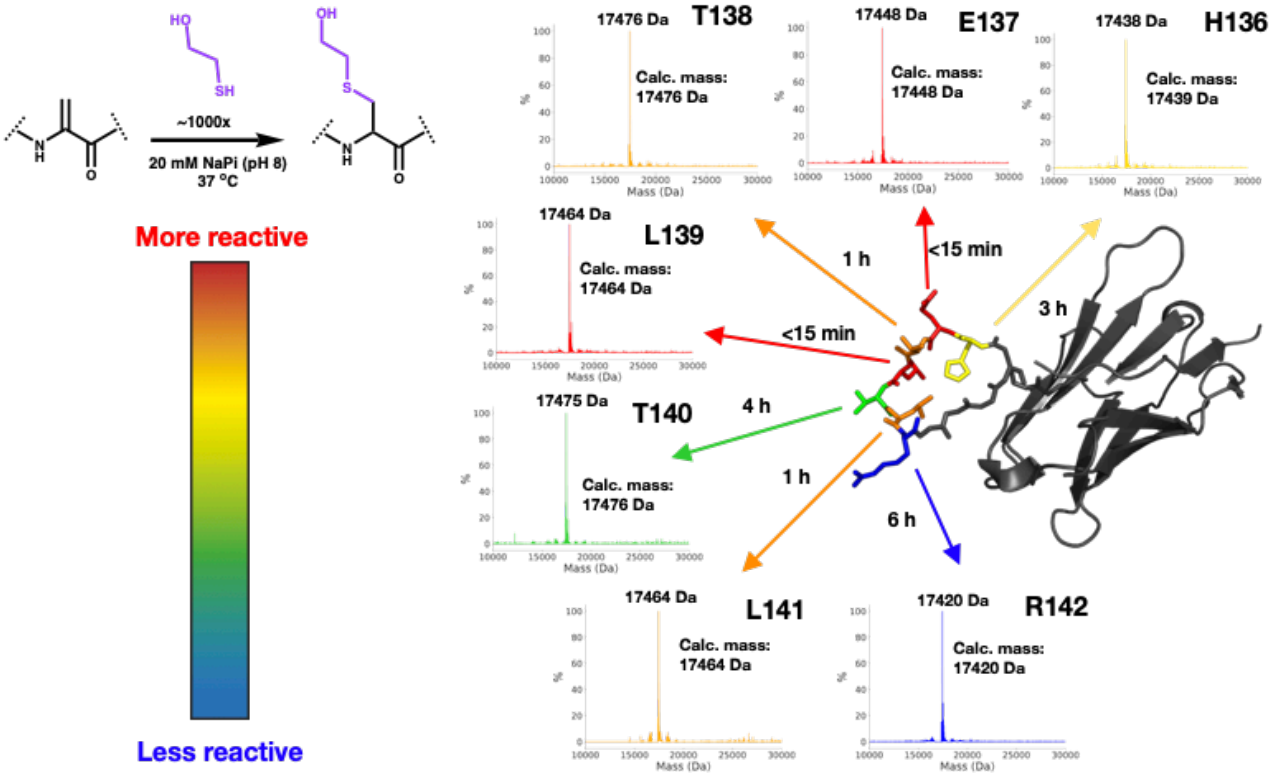
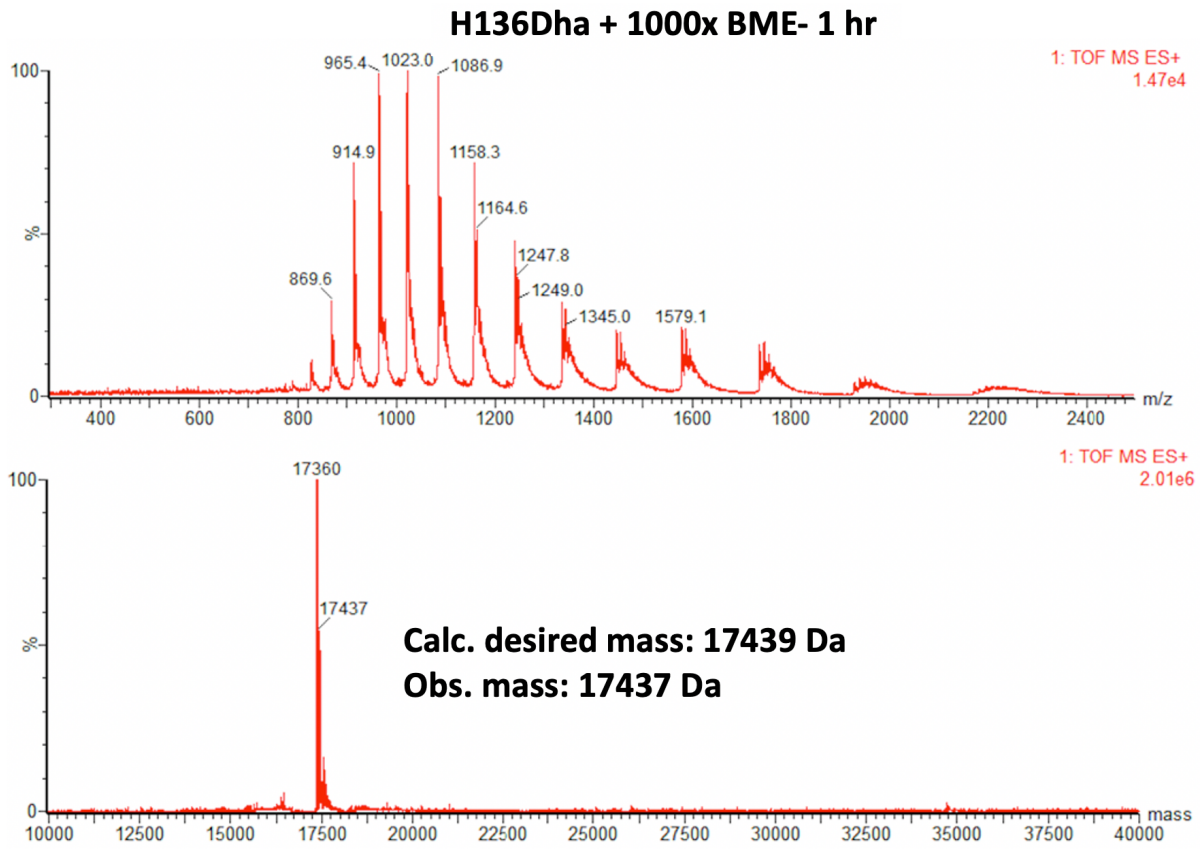
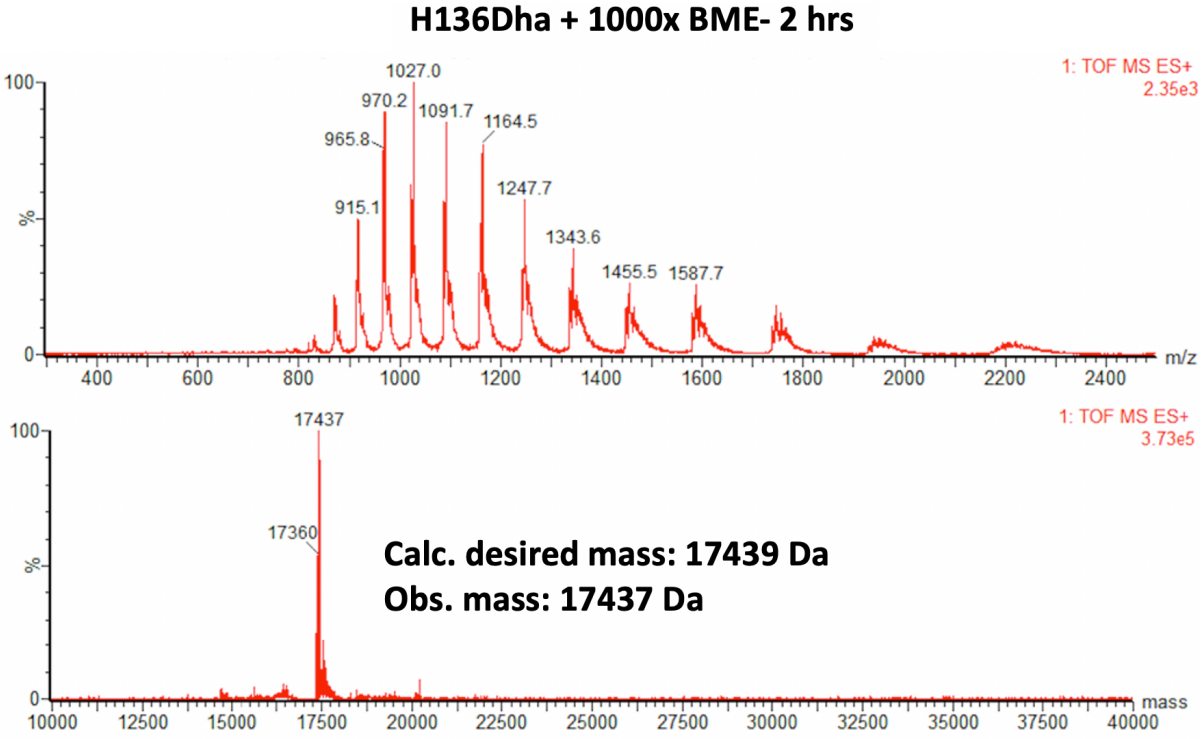


Figure 2.4. Dha reactivity map along the CDR3-loop of DesAb-Aβ$_{(3-9)}$. Time to completion for each of the Dha positions. for the full reaction time course see panels a-v below. Time points were taken by aliquoting 4 µL of the reaction mixture and diluting it with 16 µL of the reaction buffer (20 mM NaPi (pH 8)), 5 µL of the diluted samples were injected into the LC-MS.

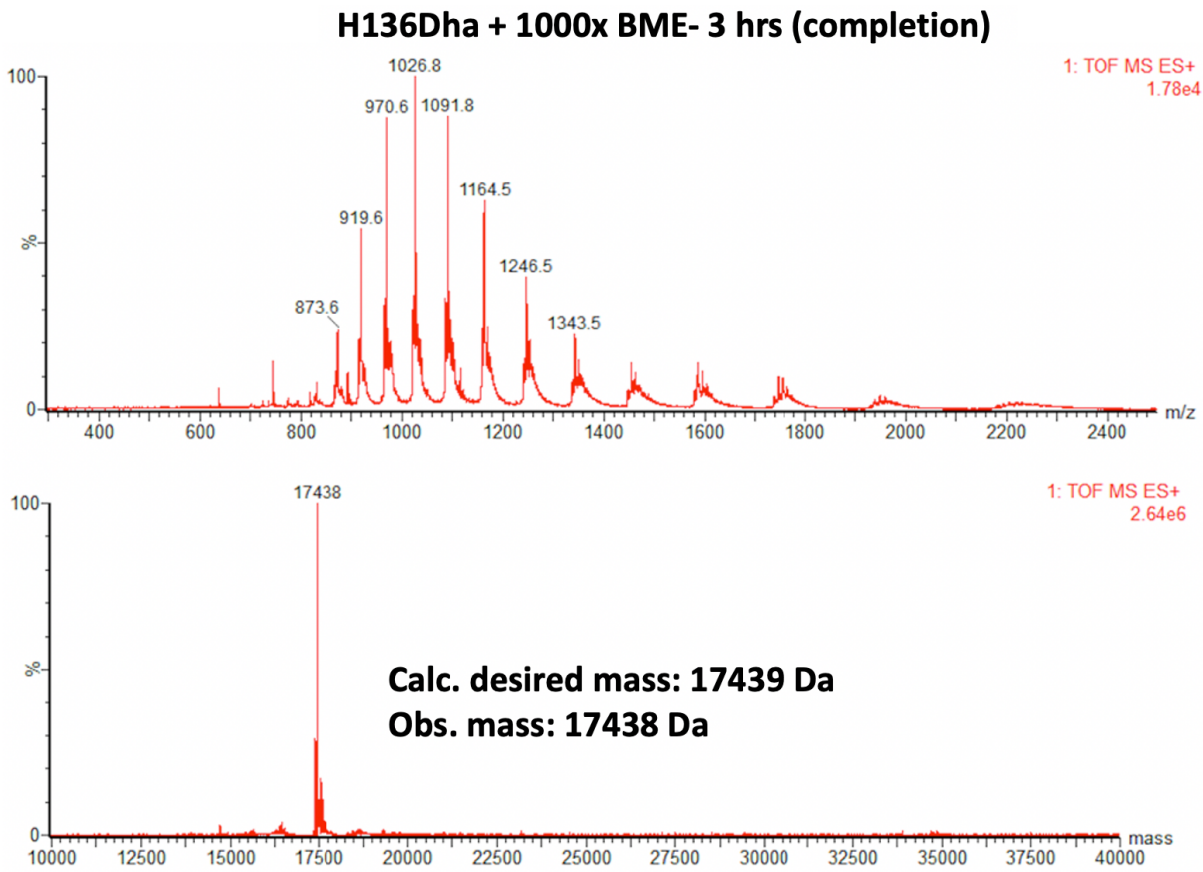
a


H136Dha + 1000x BME- 1 hr
1: TOF MS ES+
1.47e4
965.4
1023.0
1086.9
914.9
1158.3
1164.6
1247.8
869.6
1249.0
1345.0
1579.1
m/z
1: TOF MS ES+
2.01e6
17360
17437
Calc. desired mass: 17439 Da
Obs. mass: 17437 Da
mass

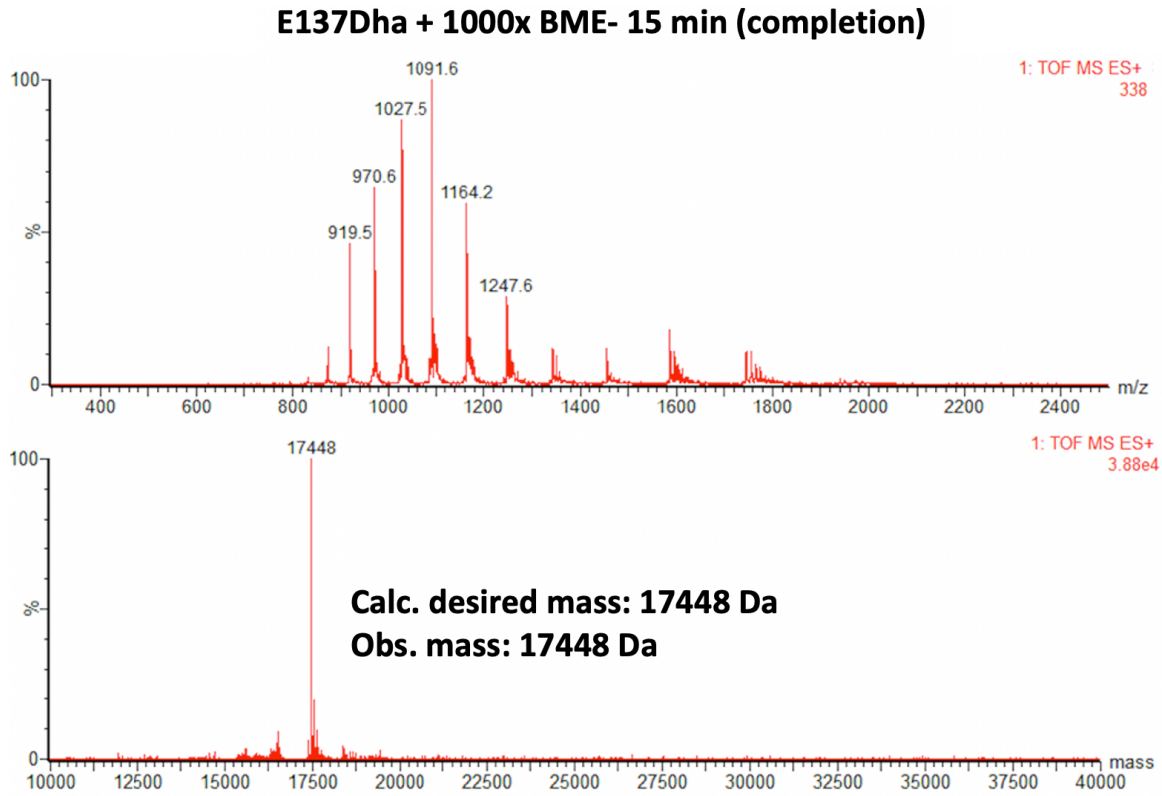
b


H136Dha + 1000x BME- 2 hrs
1: TOF MS ES+
2.35e3
1027.0
970.2
1091.7
965.8
1164.5
915.1
1247.7
1343.6
1455.5
1587.7
m/z
1: TOF MS ES+
3.73e5
17437
17360
Calc. desired mass: 17439 Da
Obs. mass: 17437 Da
mass

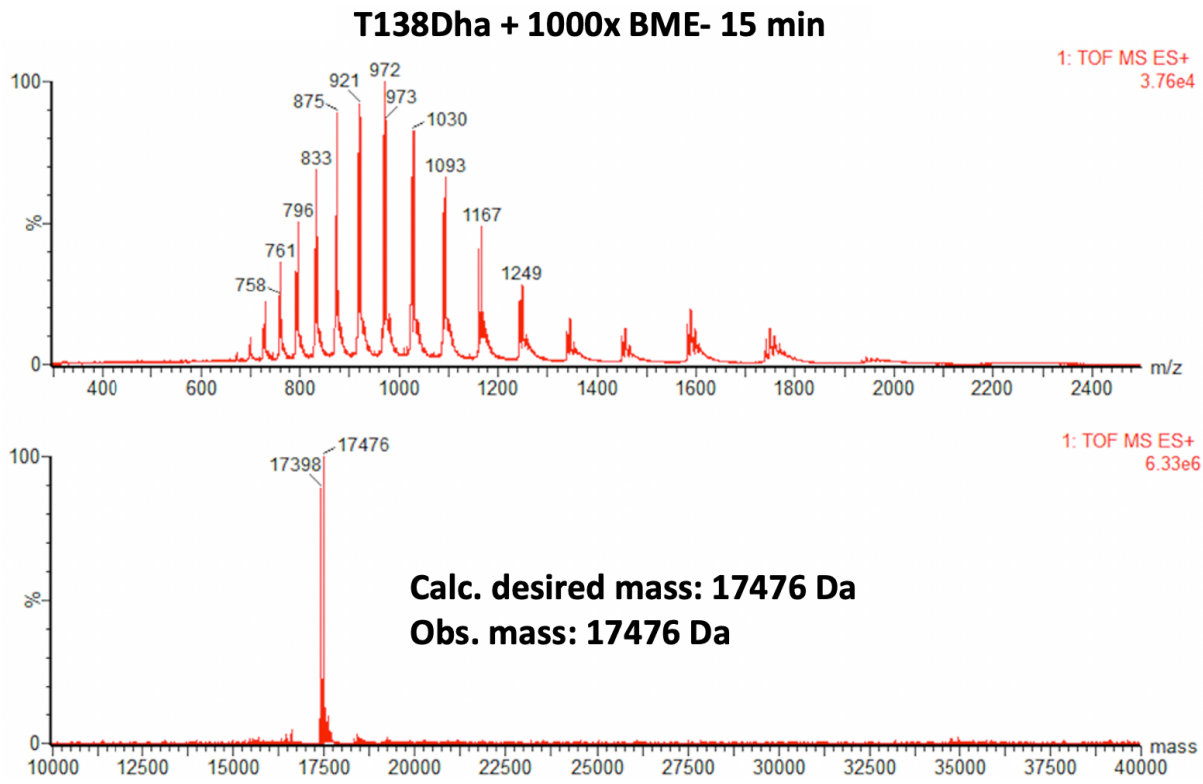
c


H136Dha + 1000x BME- 3 hrs (completion)
1: TOF MS ES+
1.78e4
1026.8
970.6
1091.8
1164.5
919.6
1246.5
873.6
1343.5
m/z
1: TOF MS ES+
2.64e6
17438
Calc. desired mass: 17439 Da
Obs. mass: 17438 Da
mass

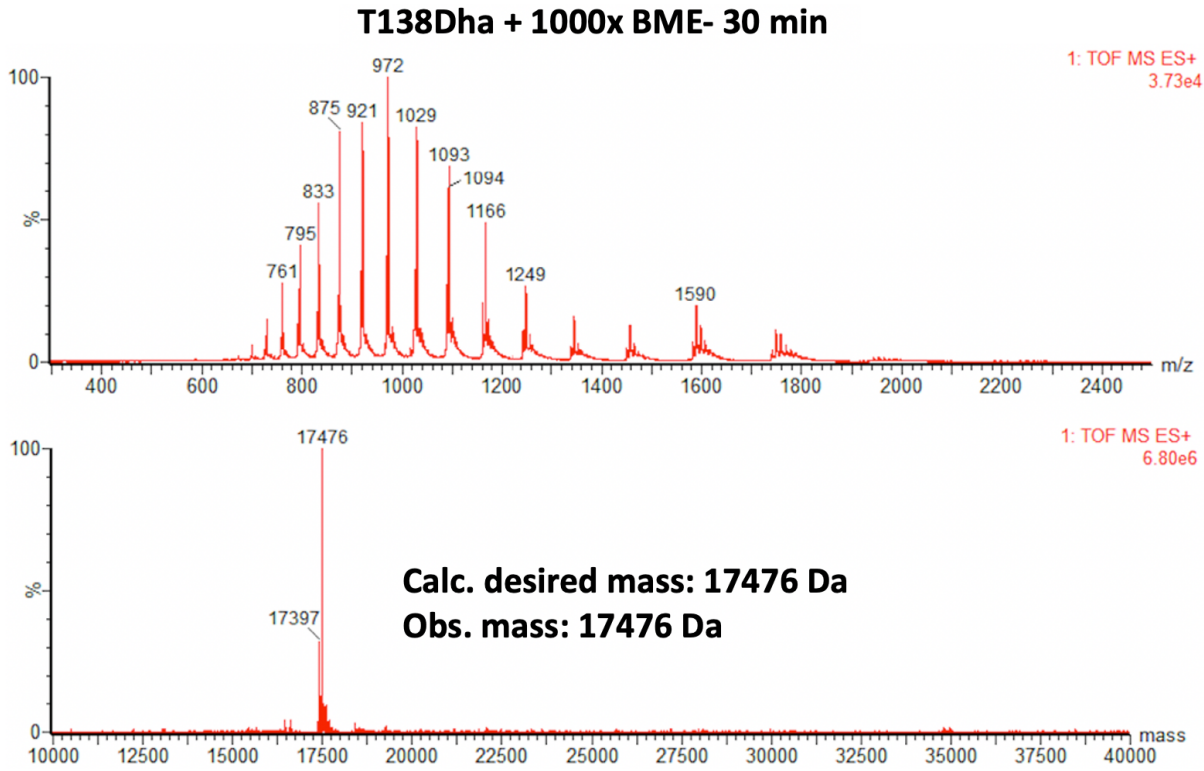
d


E137Dha + 1000x BME- 15 min (completion)
1: TOF MS ES+
338
1091.6
1027.5
970.6
1164.2
919.5
1247.6
m/z
1: TOF MS ES+
3.88e4
17448
Calc. desired mass: 17448 Da
Obs. mass: 17448 Da
mass

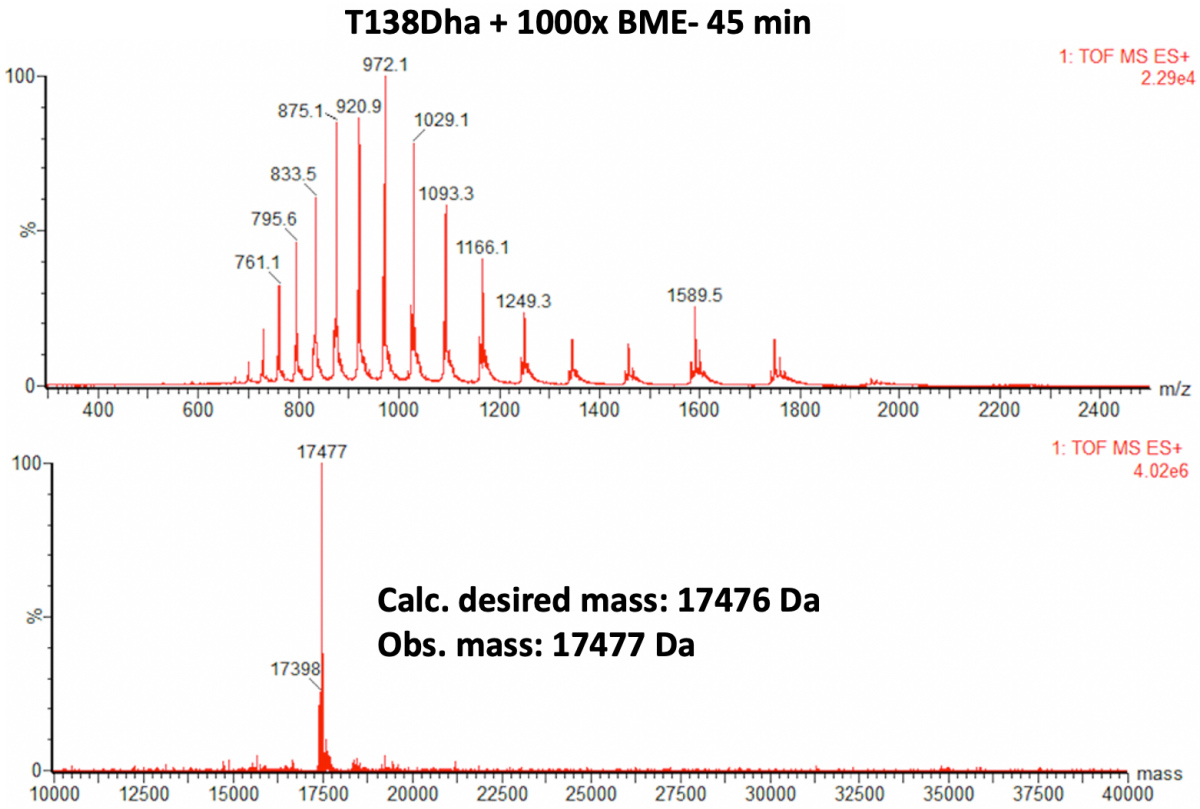
e


T138Dha + 1000x BME- 15 min
1: TOF MS ES+
3.76e4
100
%
0
758
761
796
833
875
921
972
973
1030
1093
1167
1249
m/z
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
1: TOF MS ES+
6.33e6
17398
17476
Calc. desired mass: 17476 Da
Obs. mass: 17476 Da
mass
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000

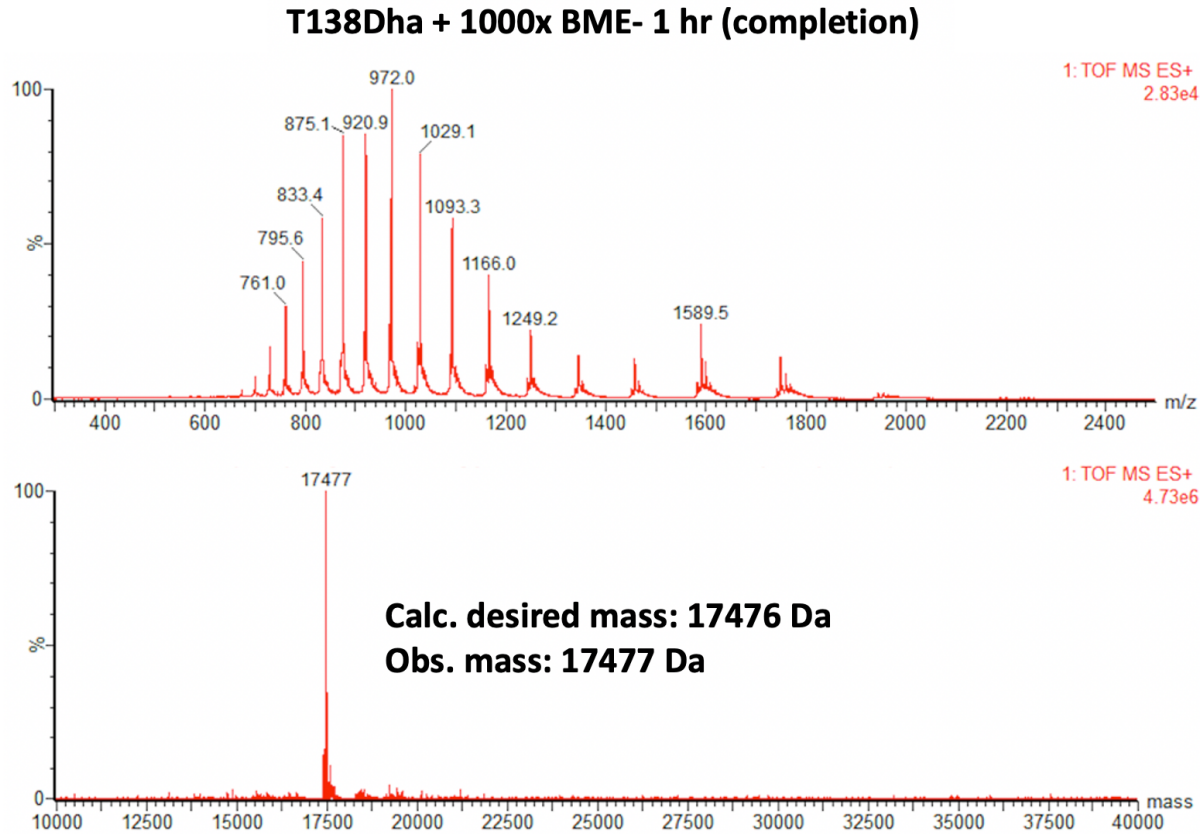
f


T138Dha + 1000x BME- 30 min
1: TOF MS ES+
3.73e4
100
%
0
761
795
833
875
921
972
1029
1093
1094
1166
1249
1590
m/z
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
1: TOF MS ES+
6.80e6
17397
17476
Calc. desired mass: 17476 Da
Obs. mass: 17476 Da
mass
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000

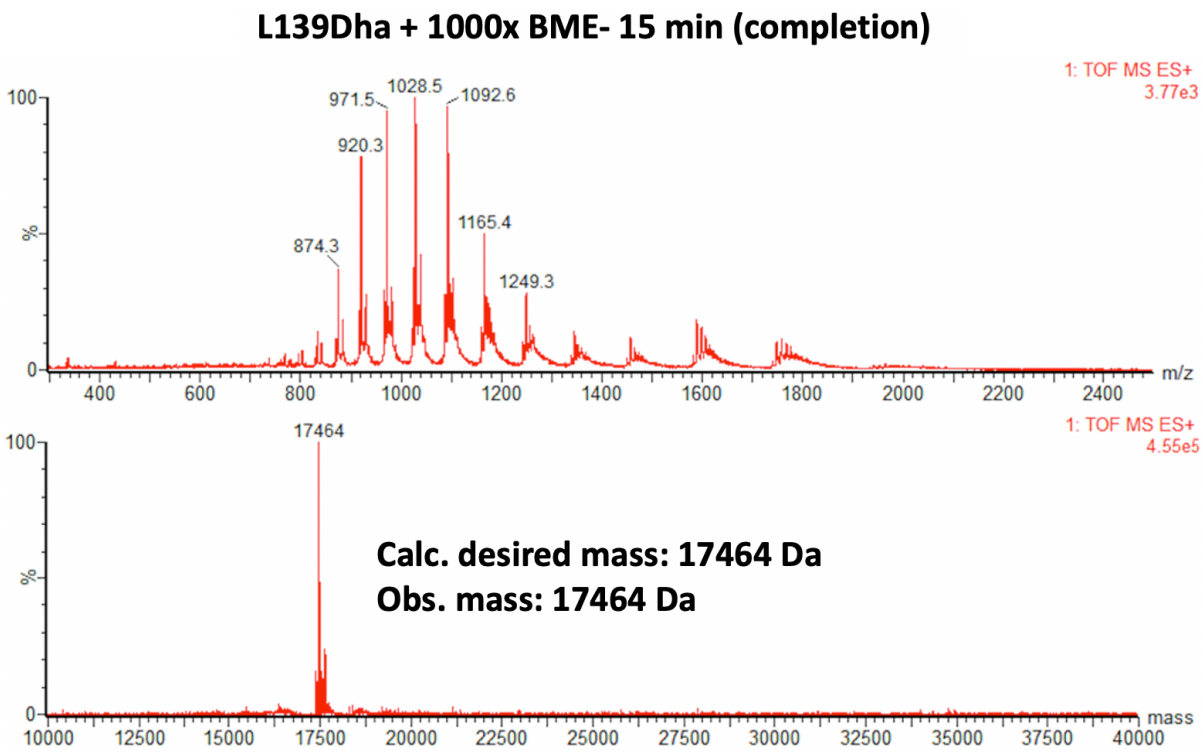
g


T138Dha + 1000x BME- 45 min
1: TOF MS ES+
2.29e4
100
%
0
972.1
875.1
920.9
1029.1
833.5
1093.3
795.6
1166.1
761.1
1249.3
1589.5
m/z
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
1: TOF MS ES+
4.02e6
17477
17398
Calc. desired mass: 17476 Da
Obs. mass: 17477 Da
mass
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000

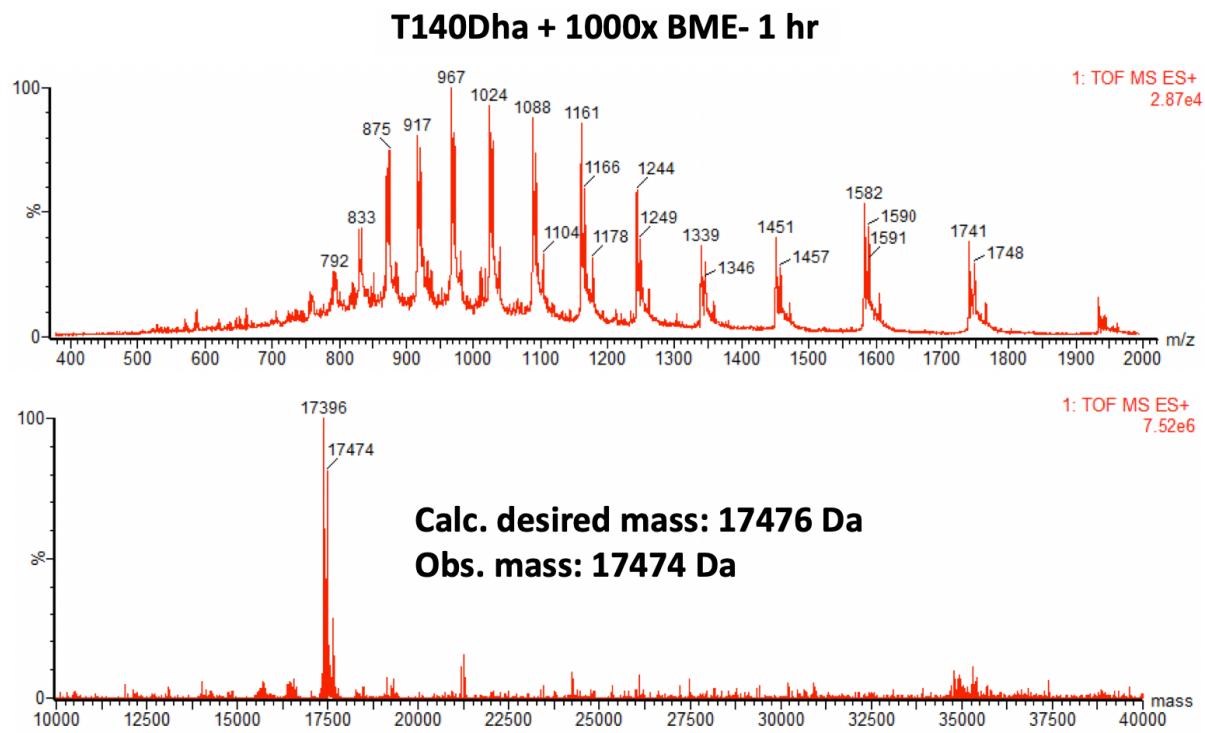
h


T138Dha + 1000x BME- 1 hr (completion)
1: TOF MS ES+
2.83e4
100
%
0
972.0
875.1
920.9
1029.1
833.4
1093.3
795.6
1166.0
761.0
1249.2
1589.5
m/z
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
1: TOF MS ES+
4.73e6
17477
Calc. desired mass: 17476 Da
Obs. mass: 17477 Da
mass
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000

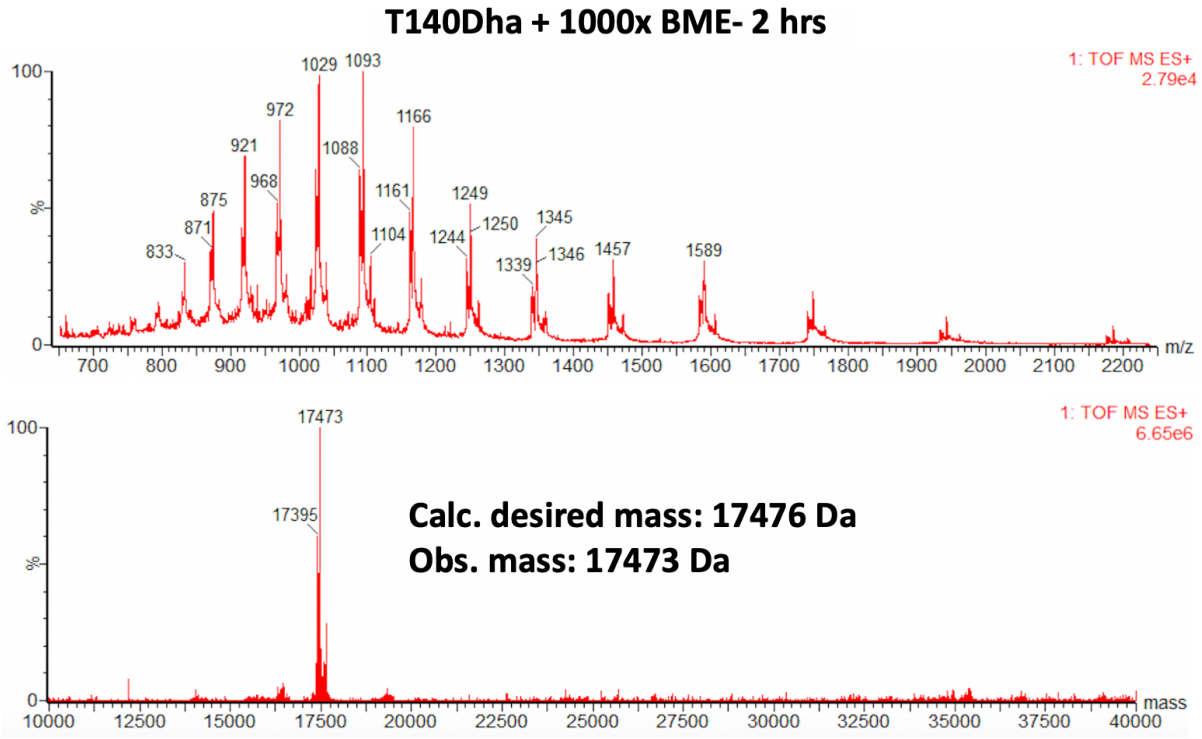
i


L139Dha + 1000x BME- 15 min (completion)
1: TOF MS ES+
3.77e3
971.5
1028.5
1092.6
920.3
1165.4
874.3
1249.3
m/z
1: TOF MS ES+
4.55e5
17464
Calc. desired mass: 17464 Da
Obs. mass: 17464 Da
mass

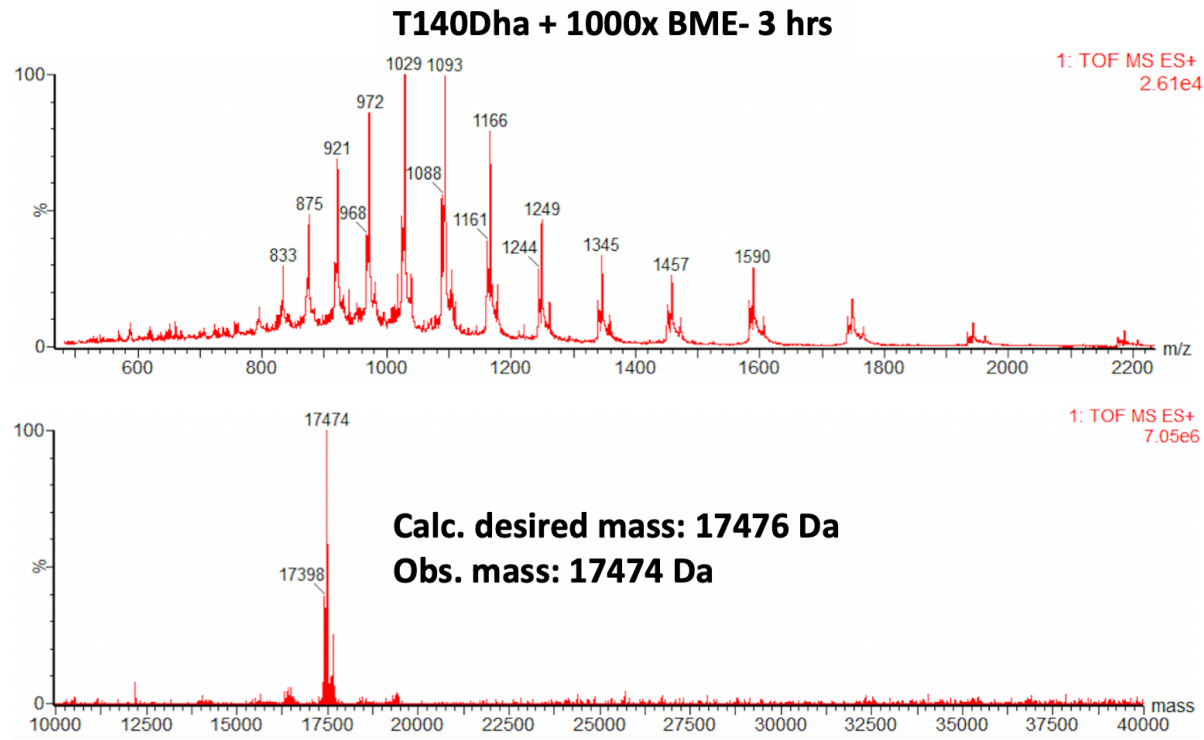
j


T140Dha + 1000x BME- 1 hr
1: TOF MS ES+
2.87e4
967
1024
1088
1161
875
917
1166
1244
833
1582
1590
1249
1104
1178
1339
1451
1591
1741
792
1346
1457
1748
m/z
1: TOF MS ES+
7.52e6
17396
17474
Calc. desired mass: 17476 Da
Obs. mass: 17474 Da
mass

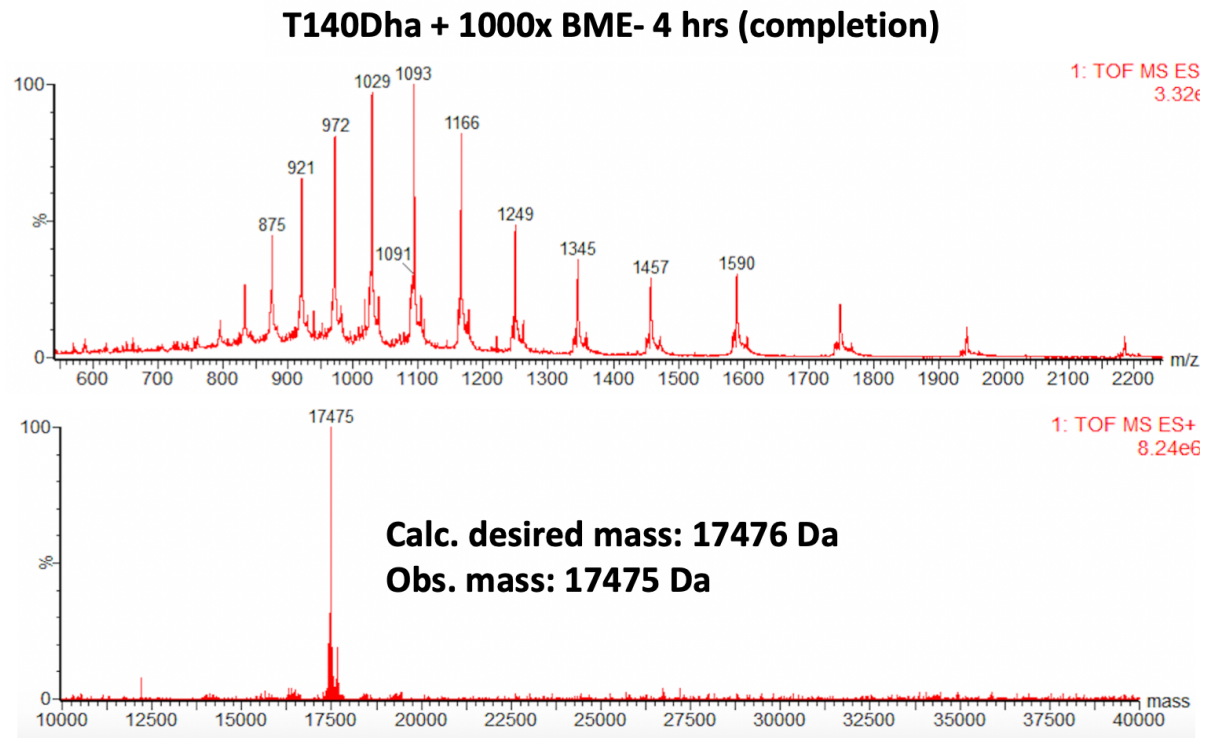
k


T140Dha + 1000x BME- 2 hrs
1: TOF MS ES+
2.79e4
1029
1093
972
1166
921
1088
968
1161
875
1249
871
1104
1250
1345
833
1244
1339
1346
1457
1589
m/z
1: TOF MS ES+
6.65e6
17473
17395
Calc. desired mass: 17476 Da
Obs. mass: 17473 Da
mass

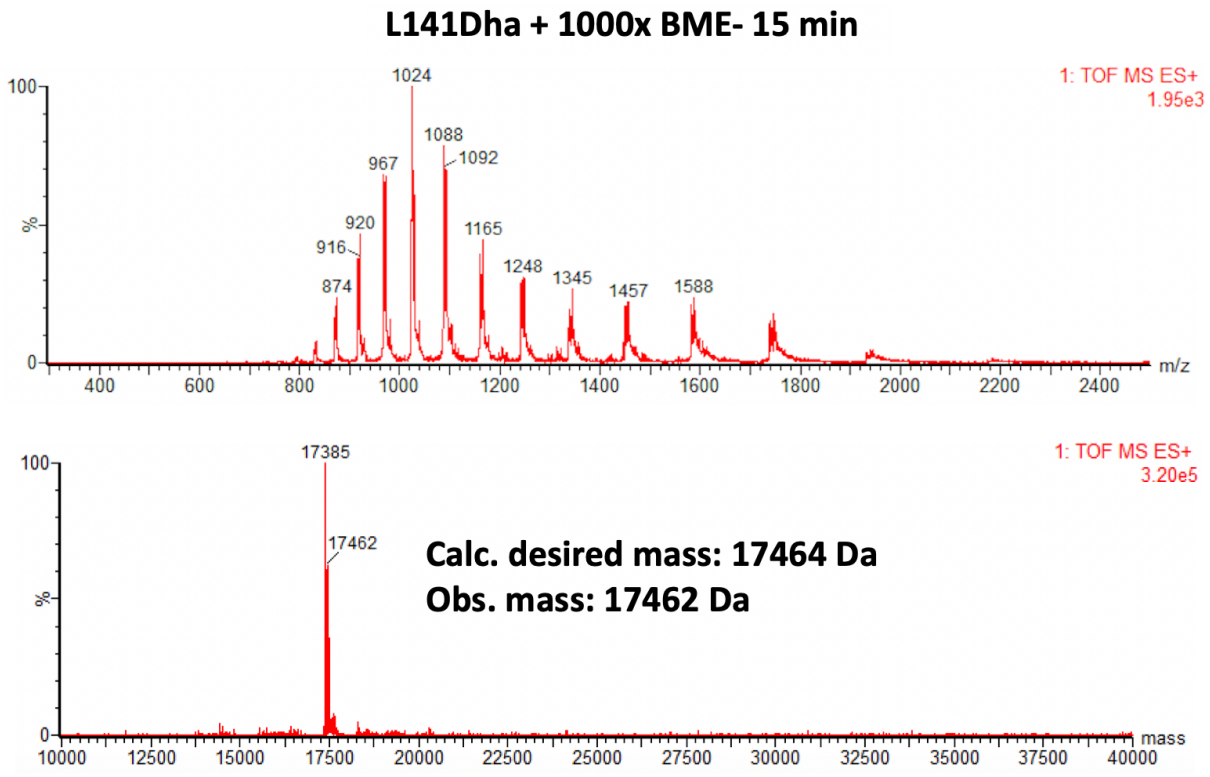
l


T140Dha + 1000x BME- 3 hrs
1: TOF MS ES+
2.61e4
1029
1093
972
1166
921
1088
875
968
1161
1249
833
1244
1345
1457
1590
m/z
1: TOF MS ES+
7.05e6
17474
Calc. desired mass: 17476 Da
17398
Obs. mass: 17474 Da
mass

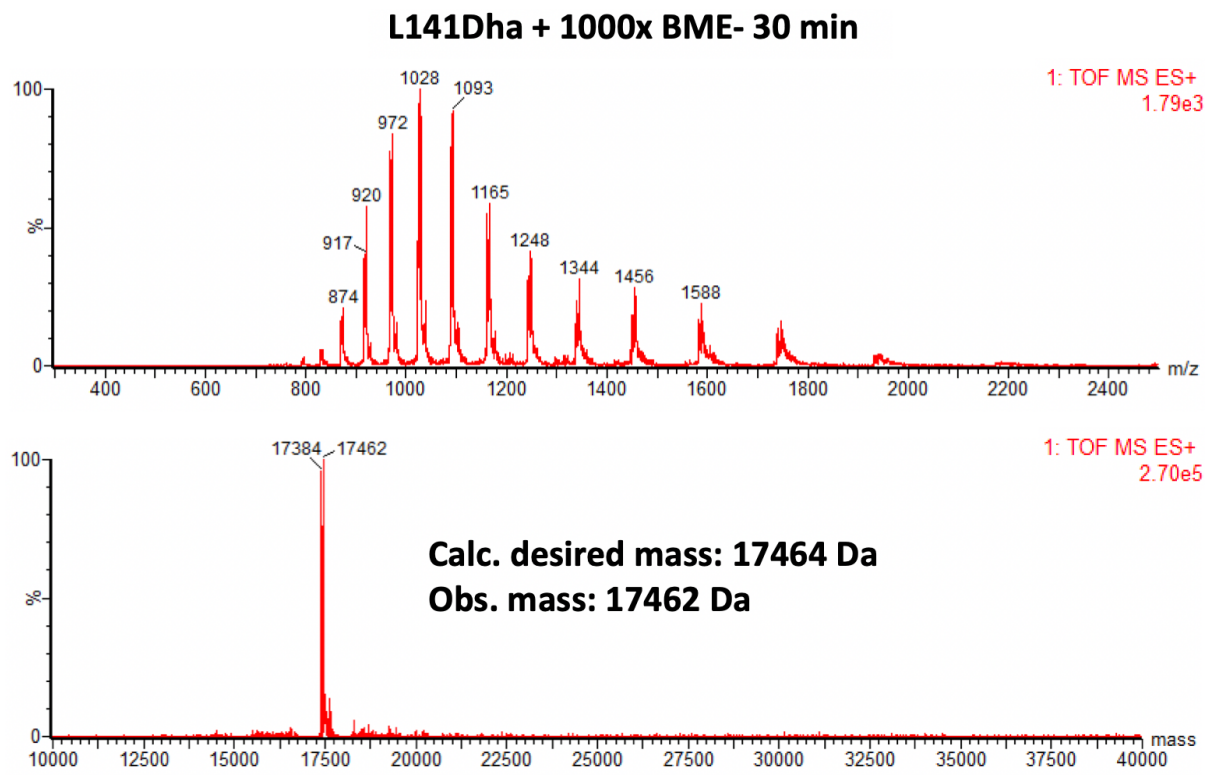
m


T140Dha + 1000x BME- 4 hrs (completion)
1: TOF MS ES
3.32e
1029
1093
972
1166
921
875
1249
1091
1345
1457
1590
m/z
1: TOF MS ES+
8.24e6
17475
Calc. desired mass: 17476 Da
Obs. mass: 17475 Da
mass

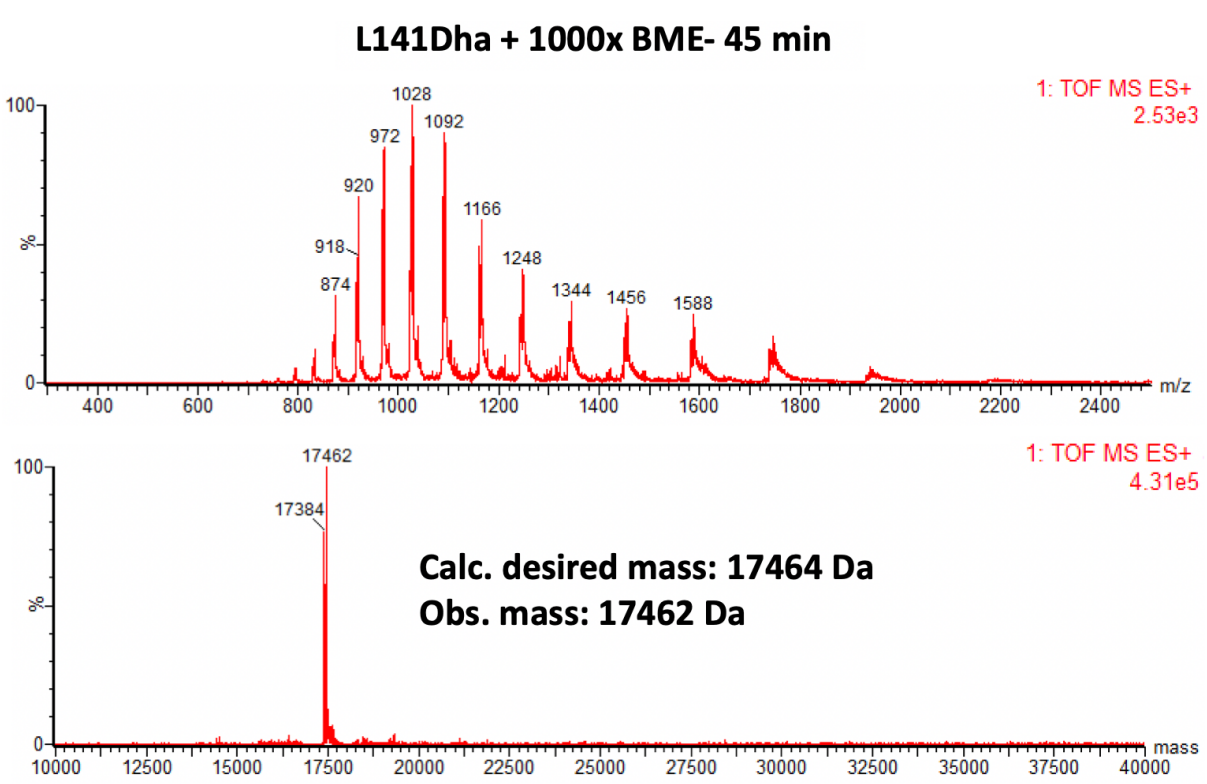
n


L141Dha + 1000x BME- 15 min
1: TOF MS ES+
1.95e3
1024
1088
1092
967
920
916
1165
874
1248
1345
1457
1588
m/z
1: TOF MS ES+
3.20e5
17385
17462
Calc. desired mass: 17464 Da
Obs. mass: 17462 Da
mass

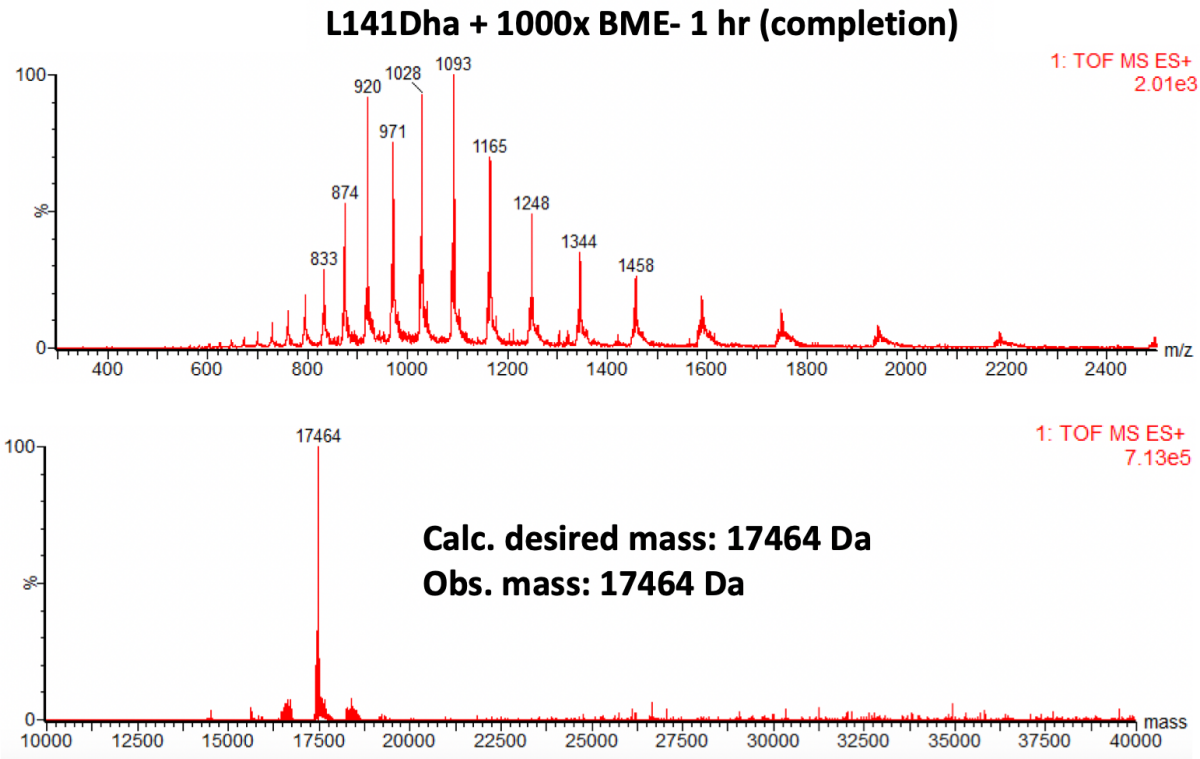
o


L141Dha + 1000x BME- 30 min
1: TOF MS ES+
1.79e3
1028
1093
972
920
1165
917
1248
1344
1456
874
1588
m/z
1: TOF MS ES+
2.70e5
17384
17462
Calc. desired mass: 17464 Da
Obs. mass: 17462 Da
mass

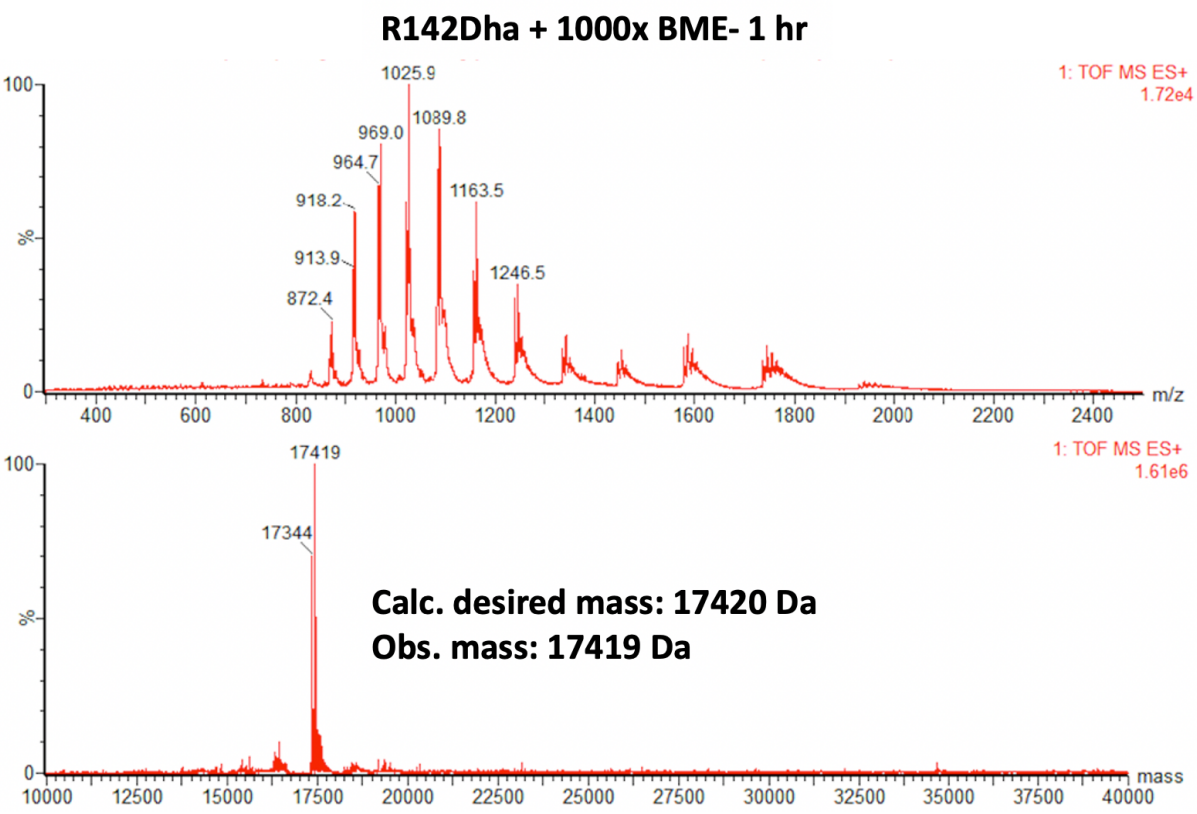
p


L141Dha + 1000x BME- 45 min
1: TOF MS ES+
2.53e3
1028
1092
972
920
1166
918
1248
874
1344
1456
1588
m/z
1: TOF MS ES+
4.31e5
17462
17384
Calc. desired mass: 17464 Da
Obs. mass: 17462 Da
mass

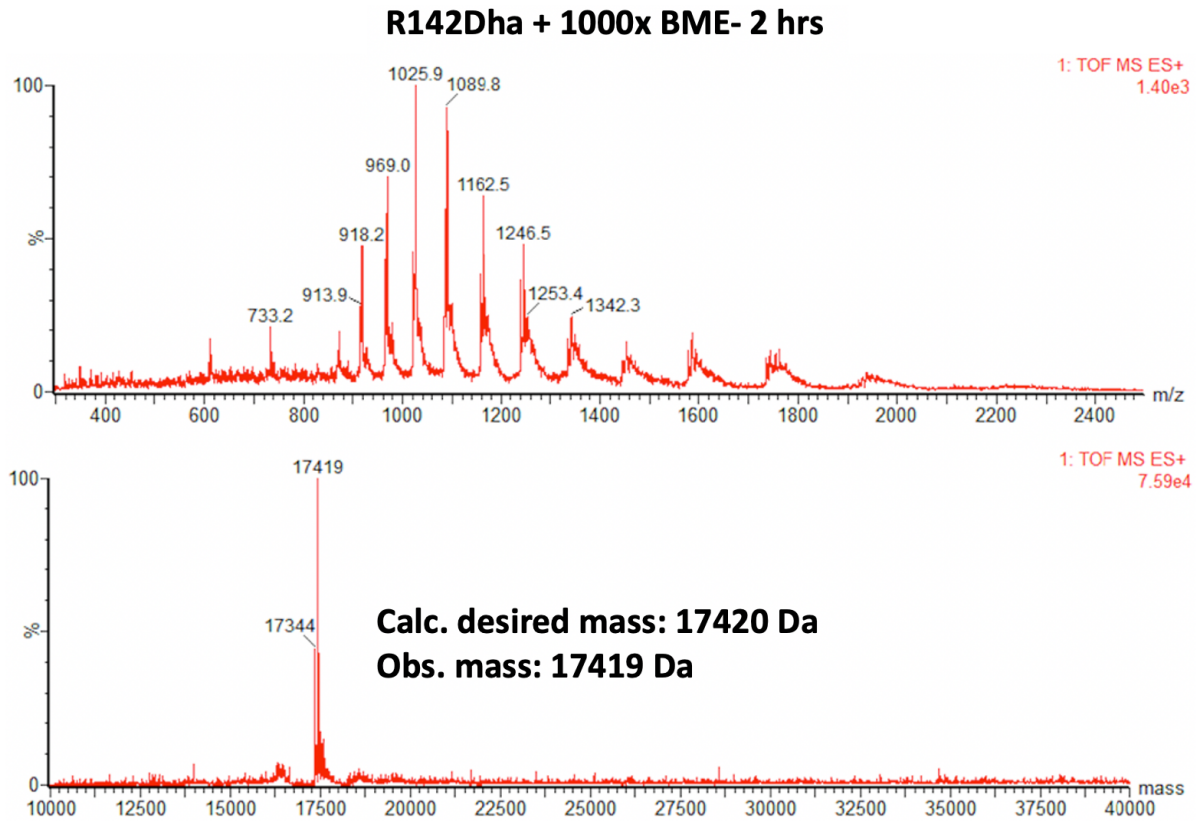
q


L141Dha + 1000x BME- 1 hr (completion)
1: TOF MS ES+
2.01e3
100
%
0
833
874
920
971
1028
1093
1165
1248
1344
1458
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
7.13e5
17464
Calc. desired mass: 17464 Da
Obs. mass: 17464 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass

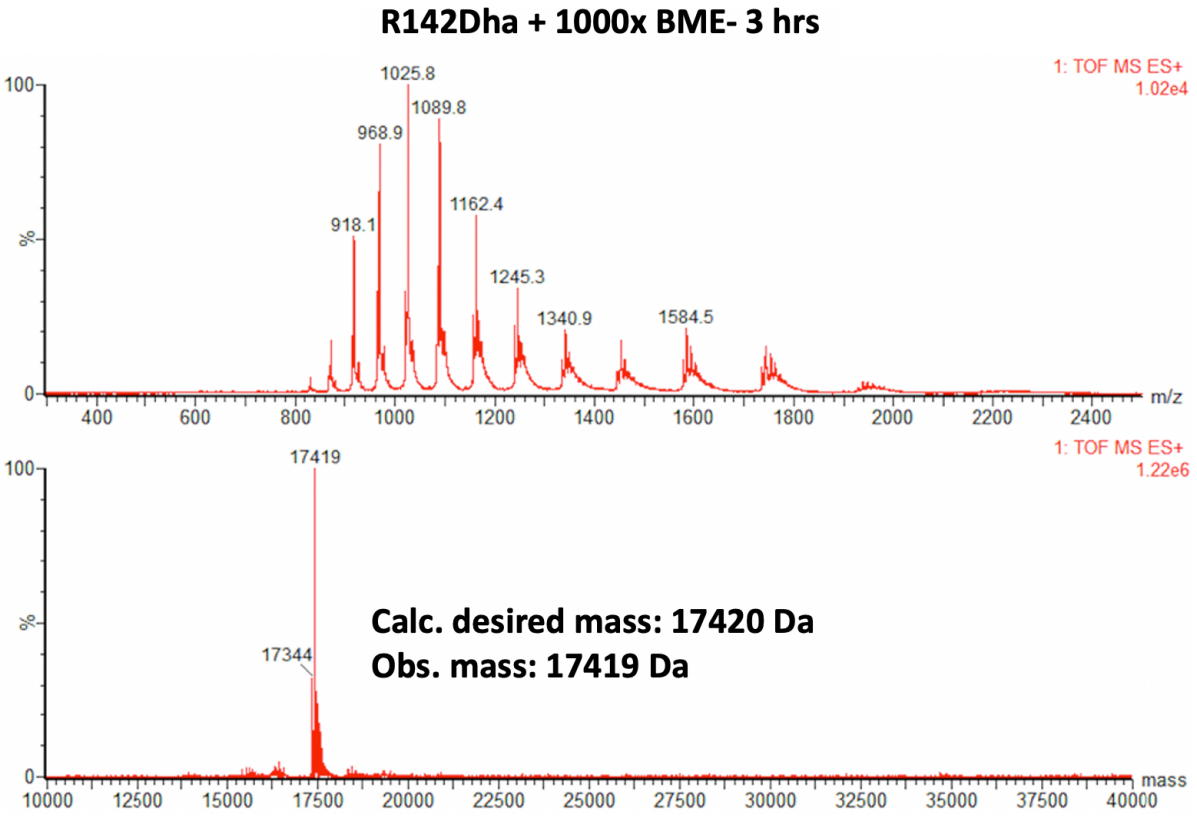
r


R142Dha + 1000x BME- 1 hr
1: TOF MS ES+
1.72e4
100
%
0
872.4
913.9
918.2
964.7
969.0
1025.9
1089.8
1163.5
1246.5
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
1.61e6
17344
17419
Calc. desired mass: 17420 Da
Obs. mass: 17419 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass

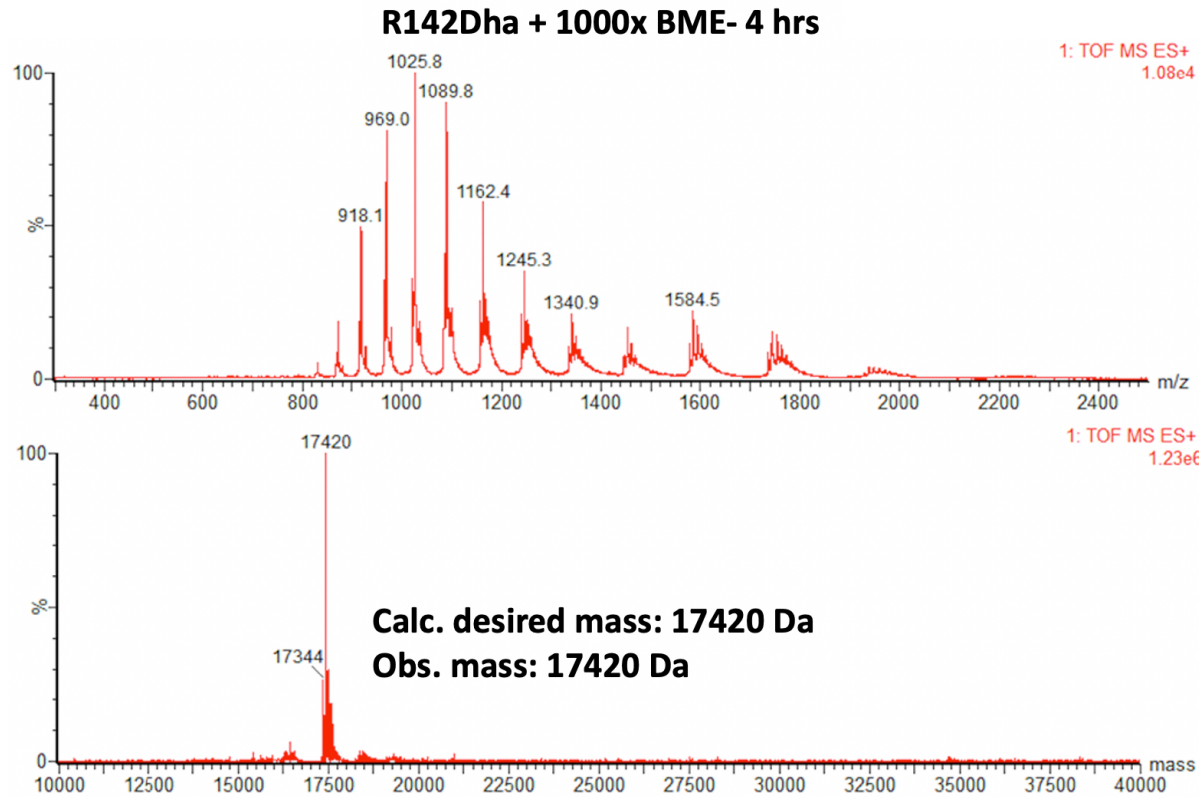
s


R142Dha + 1000x BME- 2 hrs
1: TOF MS ES+
1.40e3
1025.9
1089.8
969.0
1162.5
918.2
1246.5
913.9
1253.4
733.2
1342.3
m/z
1: TOF MS ES+
7.59e4
17419
17344
Calc. desired mass: 17420 Da
Obs. mass: 17419 Da
mass

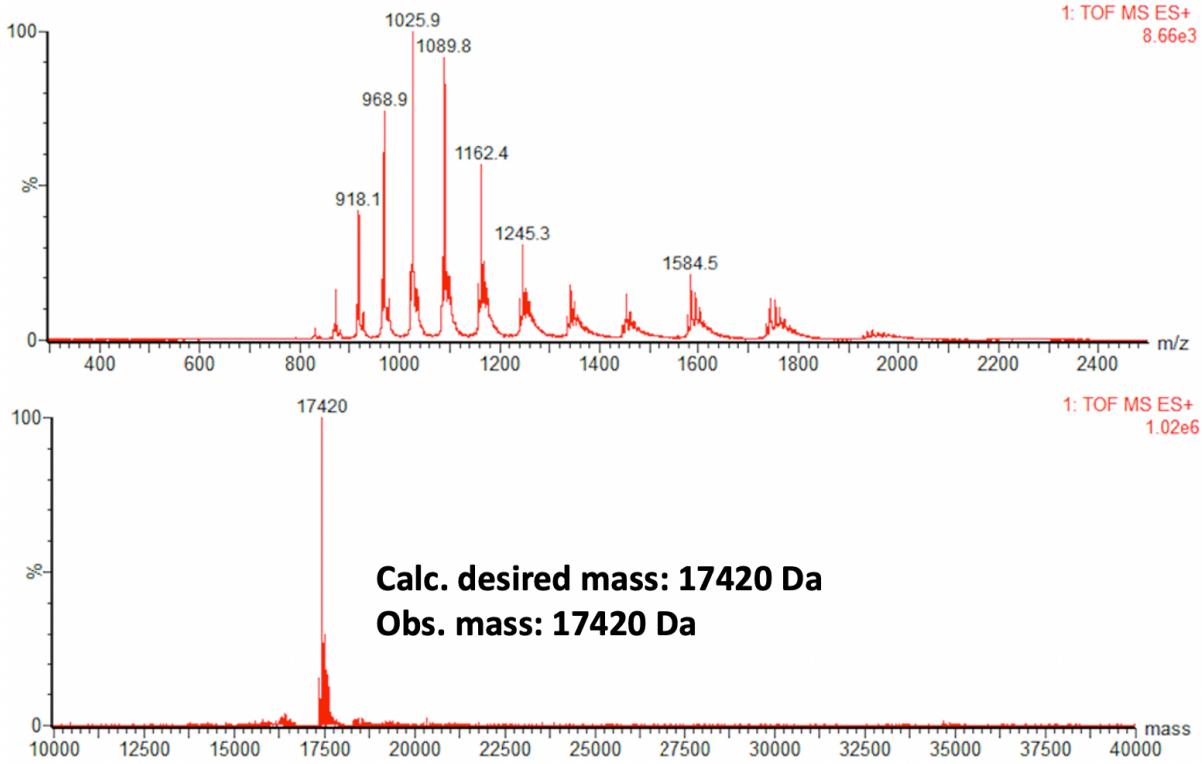
t


R142Dha + 1000x BME- 3 hrs
1: TOF MS ES+
1.02e4
1025.8
1089.8
968.9
1162.4
918.1
1245.3
1340.9
1584.5
m/z
1: TOF MS ES+
1.22e6
17419
17344
Calc. desired mass: 17420 Da
Obs. mass: 17419 Da
mass

u


R142Dha + 1000x BME- 4 hrs
1: TOF MS ES+
1.08e4
1025.8
1089.8
969.0
1162.4
918.1
1245.3
1340.9
1584.5
m/z
1: TOF MS ES+
1.23e6
17420
17344
Calc. desired mass: 17420 Da
Obs. mass: 17420 Da
mass

v


R142Dha + 1000x BME- 6 hrs (completion)
1: TOF MS ES+
8.66e3
1025.9
1089.8
968.9
1162.4
918.1
1245.3
1584.5
m/z
1: TOF MS ES+
1.02e6
17420
Calc. desired mass: 17420 Da
Obs. mass: 17420 Da
mass

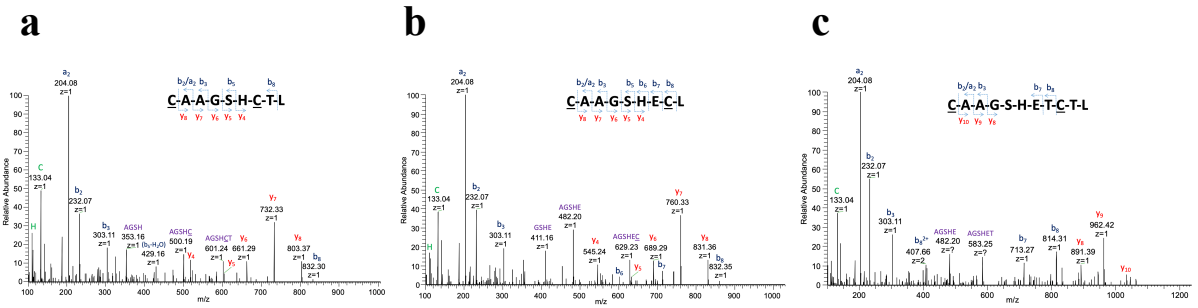


Figure 2.5. LC-MS/MS spectra confirming β-mercaptoethanol addition sites. (a) MS/MS spectrum of the *m/z* 482.2 doubly charged ion of the chymotryptic peptide CAAGSHCTL from a protein sample containing a carbamidomethyl modification at the N-terminal cysteine residue and a β-mercaptoethanol modification at the cysteine residue at position 137. **(b)** MS/MS spectrum of the *m/z* 496.2 doubly charged ion of the chymotryptic peptide CAAGSHECL from a protein sample containing a carbamidomethyl modification at the N-terminal cysteine residue and a β-mercaptoethanol modification at the cysteine residue at position 138. **(c)** MS/MS spectrum of the *m/z* 597.24 doubly charged ion of the chymotryptic peptide CAAGSHETCTL from a protein sample containing a carbamidomethyl modification at the N-terminal cysteine residue and a β-mercaptoethanol modification at the cysteine residue at position 139.

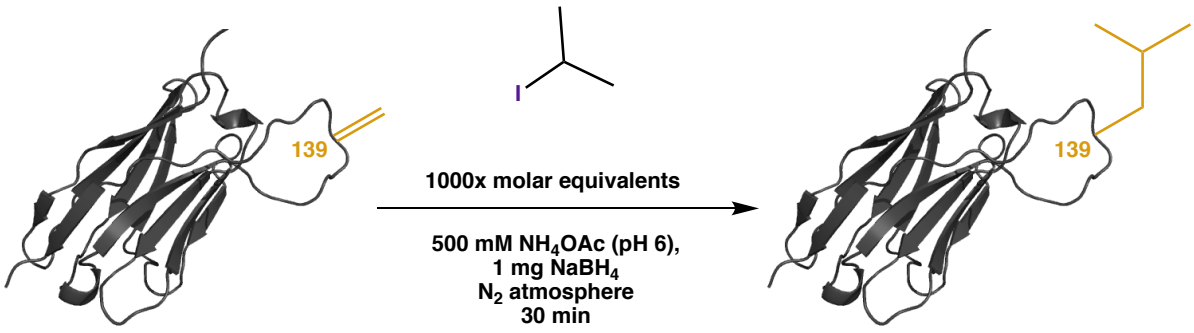
2.2.4. Reaction exploration at Dha: C(sp^3)-C(sp^3) radical mediated ligation

As mentioned previously, there are three major strategies that can be taken to selectively modify Dha: thia-Michael addition, aza-Michael addition, and free radical C(sp^3)-C(sp^3) bond formation. While the thia-Michael addition enables access to a wide array of non-canonical side chains thanks to the diverse catalogue of commercially available thiols, we wanted to explore all possible routes to enable chemical activity maturation. With the C(sp^3)-C(sp^3) reaction scheme, the precursors to the chemical mutants are organic molecules containing either iodine or bromine at the desired site of conjugation on the small molecule. For the initial development of this reaction, the original investigators relied heavily on the conjugation of 2-iodopropane to Dha, which creates an epimeric mixture of leucine at the site of modification. This led us to pursue this modification first as well to see if this scheme was amenable to modify DesAB-Aβ$_{3\text{-}9}$, especially as position L139 was one of the most reactive sites for Dha. Note that in this section of the thesis the thermally stabilising E80K mutation had not been implemented yet, so masses are slightly different than in previous and future sections.

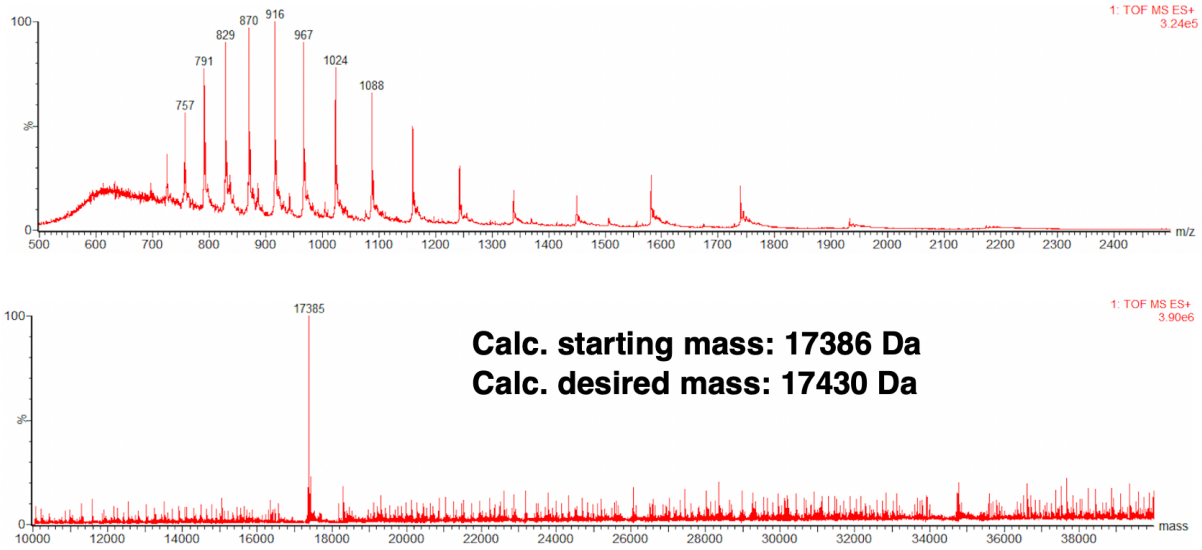
For this reaction to be possible at all, a mild free radical initiator needs to be present with the protein and halogen bearing precursor. Davis and co-workers found that, rather unexpectedly, sodium borohydride was an excellent mild source of free radicals in aqueous solution to catalyse the reaction[163]. Another important aspect of this reaction is that it should

be performed under inert conditions, as the presence of oxygen can lead to oxidative cleavage at the site of Dha. These condition guidelines led us to first attempt the modification of DesAB-$A\beta_{3-9}$ (L139Dha, K80E) with 1000 molar equivalents of 2-iodopropane, with sodium borohydride as the radical source, and in a nitrogen glovebox after degassing overnight (**Figure 2.6a**). In the morning after degassing, the protein solution was topped up with buffer to compensate for any evaporation, 2-iodopropane was added to the solution, and the reaction mixture was pipetted into a new tube containing 1 mg of $NaBH_4$. The reaction was allowed to proceed for 30 min with periodic shaking. A large amount of effervescence was always observed. However, these conditions did not lead to modification at L139Dha (**Figure 2.6b**). While L139Dha was readily modified by β-mercaptoethanol, we thought that perhaps the reactivity trend for thia-Michael additions did not completely correlate with reactivity towards free-radical addition. Therefore, we attempted the same reaction again with the addition of 6 M GndHCl to completely unfold the protein and encourage modification. However, this only led to observing a partial conversion to the undesired dialkylated product (**Figure 2.6c**). To see if the cause of these poor results was the creation of radicals by $NaBH_4$ we performed the reaction under denaturing conditions with L139Dha in open air. If there is an efficient generation of radicals nearly all of the protein should be cleaved at Dha due to the presence of oxygen. However, when this was performed only partial cleavage was observed (**Figure 2.6d**). This result led us to conclude that, for unknown reasons, $NaBH_4$ is not as efficient a radical initiator in our hands and we sought to pursue another initiator.

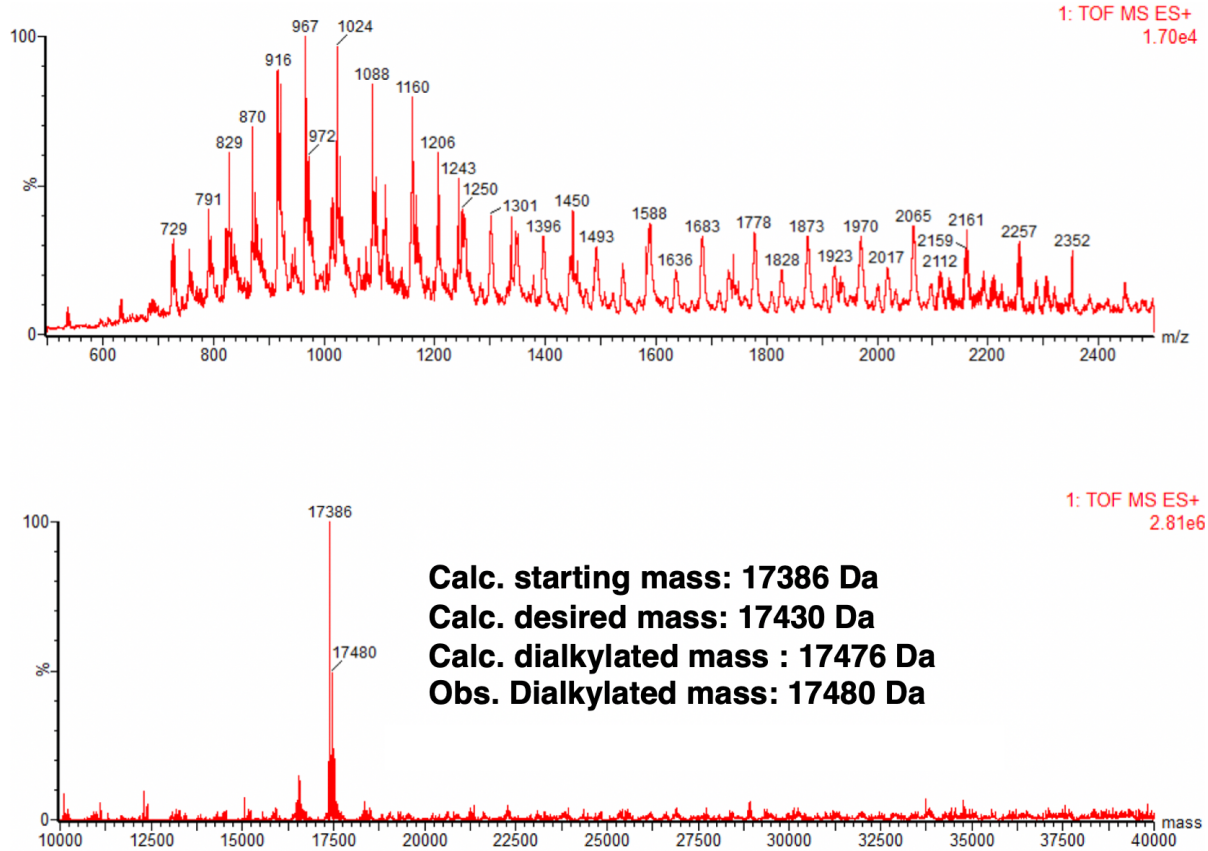
a



b


1: TOF MS ES+
3.24e5
100
%
0
757
791
829
870
916
967
1024
1088
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400
m/z
1: TOF MS ES+
3.90e6
17385
Calc. starting mass: 17386 Da
Calc. desired mass: 17430 Da
10000 12000 14000 16000 18000 20000 22000 24000 26000 28000 30000 32000 34000 36000 38000
mass

c


1: TOF MS ES+
1.70e4
100
%
0
729
791
829
870
916
967
972
1024
1088
1160
1206
1243
1250
1301
1396
1450
1493
1588
1636
1683
1778
1828
1873
1923
1970
2017
2065
2112
2159
2161
2257
2352
600 800 1000 1200 1400 1600 1800 2000 2200 2400
m/z
1: TOF MS ES+
2.81e6
17386
17480
Calc. starting mass: 17386 Da
Calc. desired mass: 17430 Da
Calc. dialkylated mass : 17476 Da
Obs. Dialkylated mass: 17480 Da
10000 12500 15000 17500 20000 22500 25000 27500 30000 32500 35000 37500 40000
mass

d

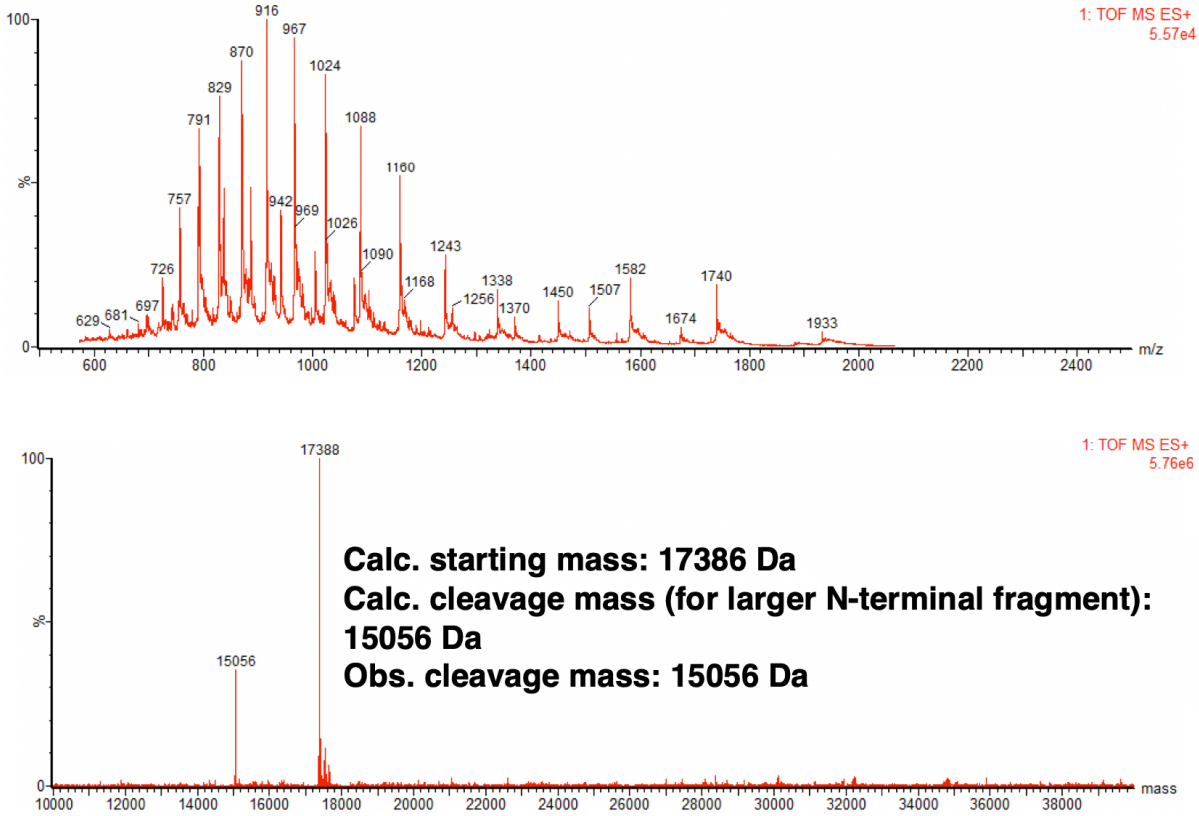


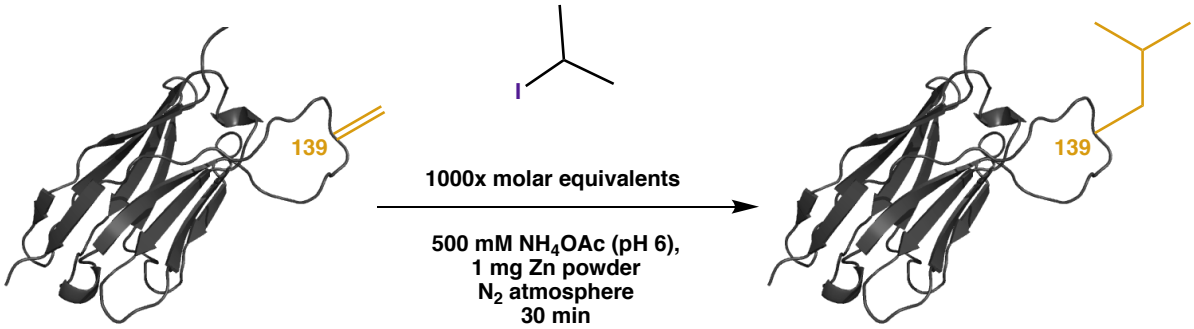
Figure 2.6. Sodium borohydride mediated free radical additions. (a) Reaction scheme for modifying DesAB-Aβ$_{3-9}$ (L139Dha, K80E) with 2-iodopropane. (b) LC-MS spectrum post-reaction under native conditions showing no modification occurred. (c) LC-MS spectrum post-reaction under denaturing conditions showing partial dialkylation. (d) LC-MS of the reaction under denaturing conditions and in the open air showing only partial cleavage.

With the negative results from using $NaBH_4$, we turned to attempting Zinc powder as the free radical initiator. Zn has been used in this reaction context before[165], the motivation for pursuing $NaBH_4$ initially was due to the safety hazards associated with powdered Zn, which is highly flammable. The same modification with 2-iodopropane was attempted under identical inert conditions as with $NaBH_4$, except that 1 mg of Zn powder was added as the initiator (**Figure 2.7a**). Under native conditions and in an N_2 atmosphere we observed a partial conversion to the desired product (**Figure 2.7b**). Encouraged by this result, we again denatured the protein in the presence of 6 M GndHCl to remove any potential structural hinderances to modification. To our dismay, this only increased the level of modification marginally (~20% to ~50% of dominant peak, **Figure 2.7c**). We again investigated the efficiency of free-radical generation by performing the reaction under denaturing conditions and in the open-air (**Figure 2.7d**). Under these conditions with the Zn powder we observed full cleavage of the protein and

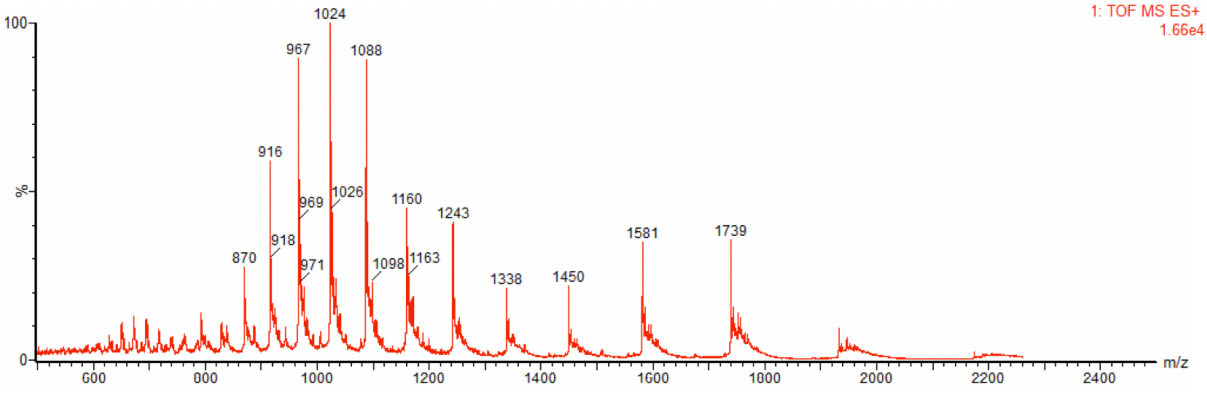
an unexpected further mass shift, possibly from oxidation of a residue during the reaction due to the presence of excess O• in the reaction. From this result it is clear that Zn powder is an efficient source for free radicals, and there is some other unknown factor that is discouraging efficient modification at this site.

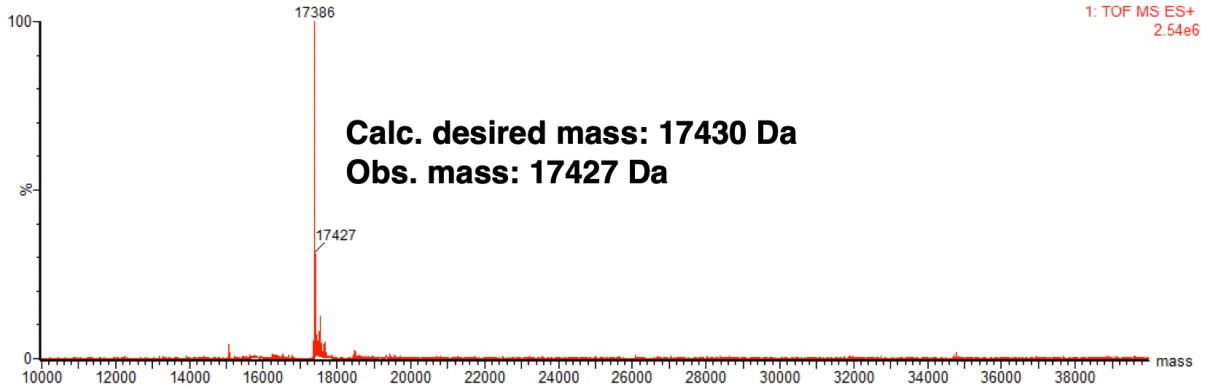
These discouraging results for the C-C bond forming reaction using two different free radical initiators were unexpected given the reported ease and wide applicability of this reaction, according to the inventors[165]. Our difficult experience with this reaction coupled with the ease of simply using thiols led us to cease pursuing this route of modification.

a

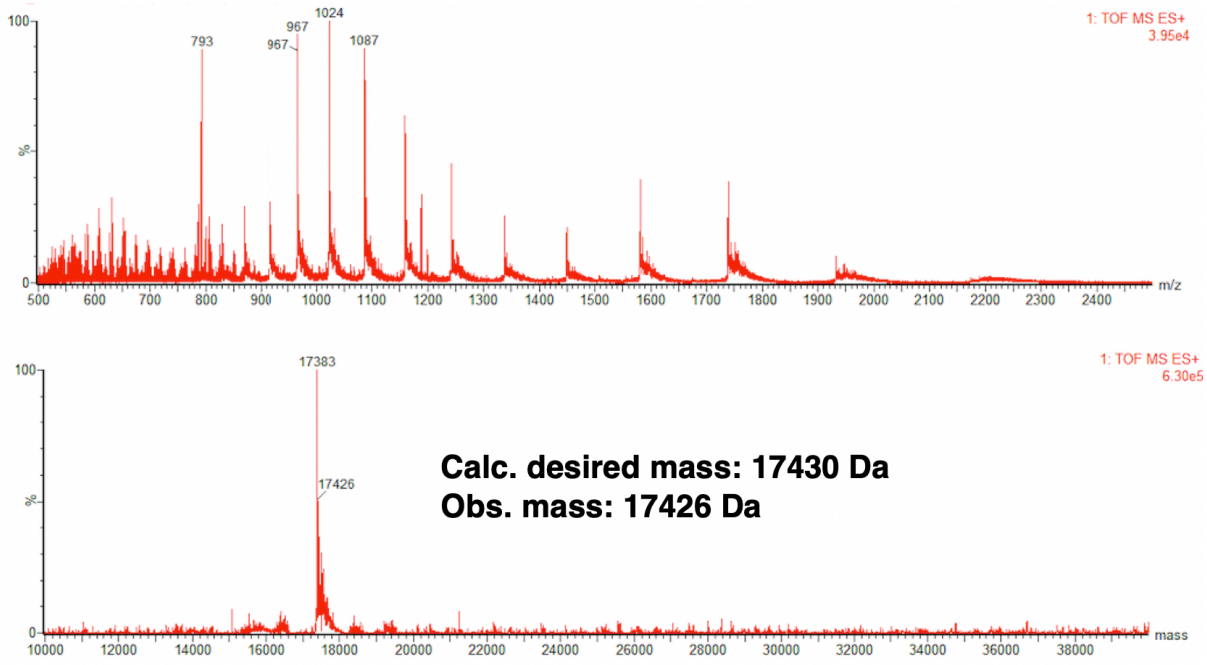


b





c



d

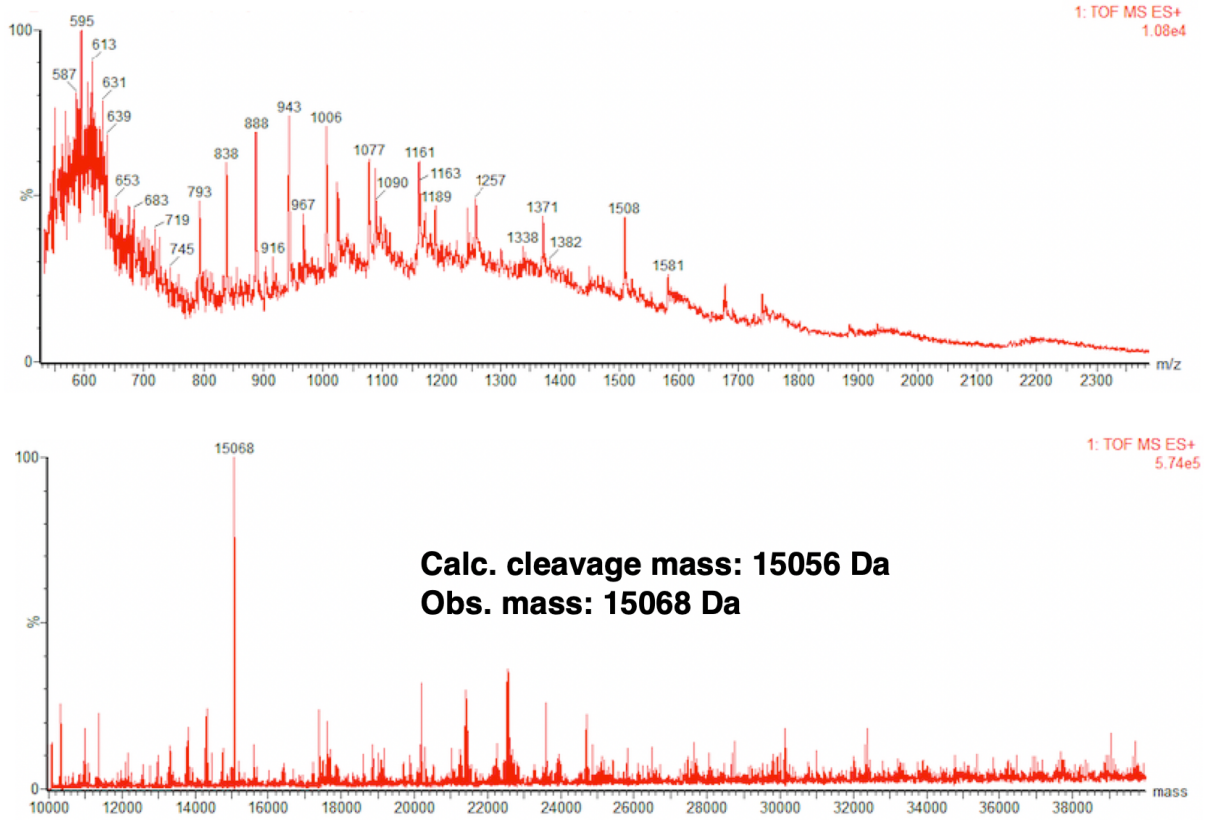


Figure 2.7. Zinc powder mediated free radical additions. (a) Reaction scheme for modifying DesAB-A$\beta_{3\text{-}9}$ (L139Dha, K80E) with 2-iodopropane. (b) LC-MS spectrum post-reaction under native conditions showing slight modification occurred. (c) LC-MS spectrum post-reaction under denaturing conditions showing similar partial alkylation as to native conditions. (d) LC-

MS of the reaction under denaturing conditions and in the open-air showing cleavage and potentially further oxidation.

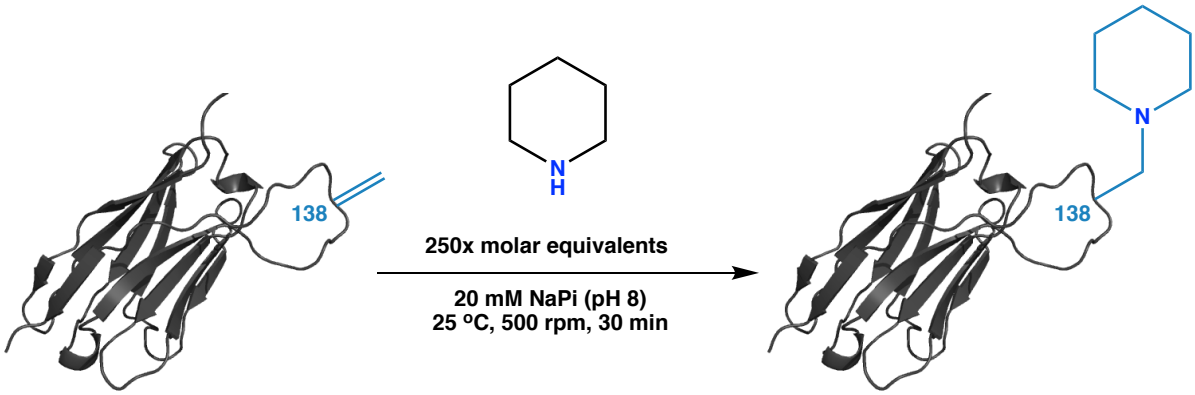
2.2.5. Reaction exploration at Dha: Aza-Michael addition

While the C-C bond forming reaction had disappointing results, we still sought to pursue another possible route of modification other than thiols. A couple more recent studies have demonstrated that Dha can be efficiently modified with secondary amines, hydrazines, and hydroxylamines[161,162]. This reaction piqued our interest in particular as this allows modification by piperidine, which is easily derivatised and one of the most common heterocycles in small molecule drugs[166]. Piperidine had previously been used to modify a camelid nanobody, so we initially pursued conditions used in that report[162]. Modification was initially pursued for DesAB-A$\beta_{3\text{-}9}$ (T138Dha, K80E), as the original threonine residue was polar and the resulting derivatives of piperidine would be as well. Performing the reaction with 250 molar equivalents and at room temperature yielded a homogenously modified product in 30 mins (**Figure 2.8b**). However, when the structural integrity was investigated by CD it was observed that the reaction had denatured the protein (**Figure 2.8d**). With the belief that the denaturation was caused by the high pH of the reaction brought about by the excess piperidine, the reaction with only 100 molar equivalents of over 1 hour was attempted. While this led to homogenous conjugation (**Figure 2.8c**), once again the CD revealed that the protein was denatured (**Figure 2.8d**).

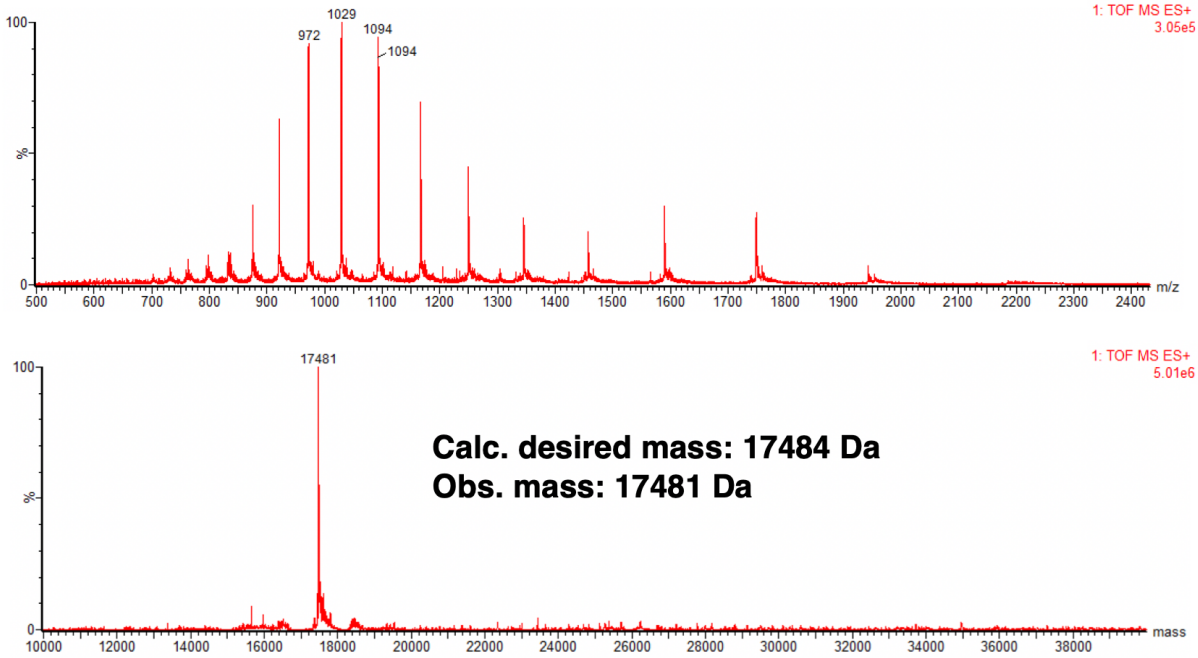
Many procedures exist to refold denatured proteins, and we were curious if we could rescue this piperidine mutant through such a protocol. 6 M GndHCl was added to the modified product after reaction with 100 molar equivalents of piperidine to completely denature the protein. After sitting on ice for 30 minutes, the sample was injected onto a Superdex 75 10/300 Increase SEC column equilibrated in 20 mM NaP_i buffer. This was done with the hope that the individual proteins would be able to safely refold on the column while the GndHCl is gradually diluted away as the run progresses. Checking the protein by CD after elution revealed that this refolding protocol worked, producing a profile largely resembling that of the starting scaffold (**Figure 2.8d**).

While the refolding protocol was successful, the extra labour needed makes this a rather cumbersome approach to conduct the side-chain exploration. With this in mind, we sought to solely pursue the thiol-based approach for Dha modification. However, with this refolding protocol in hand the aza-Michael route can now be pursued in cases where it is desired.

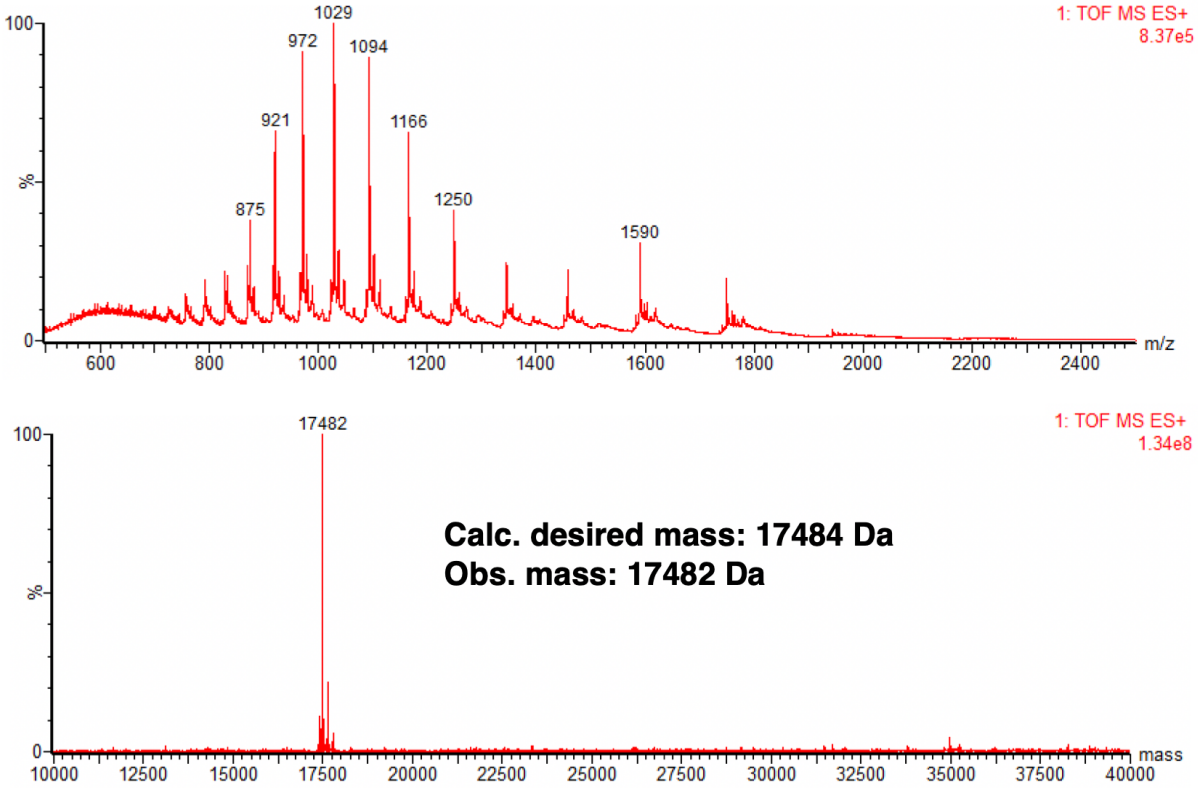
a


138
N
H
250x molar equivalents
20 mM NaPi (pH 8)
25 °C, 500 rpm, 30 min
N
138

b


1: TOF MS ES+
3.05e5
100
%
0
1029
972
1094
1094
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400
m/z
1: TOF MS ES+
5.01e6
17481
Calc. desired mass: 17484 Da
Obs. mass: 17481 Da
10000 12000 14000 16000 18000 20000 22000 24000 26000 28000 30000 32000 34000 36000 38000
mass

c



d

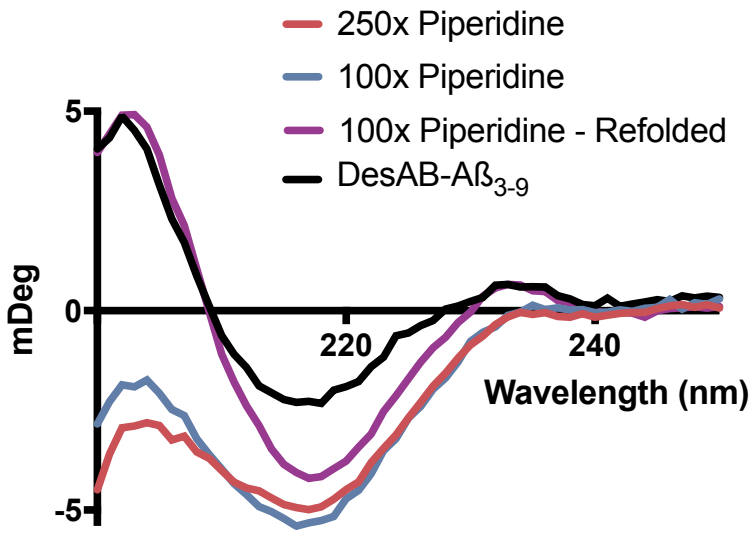
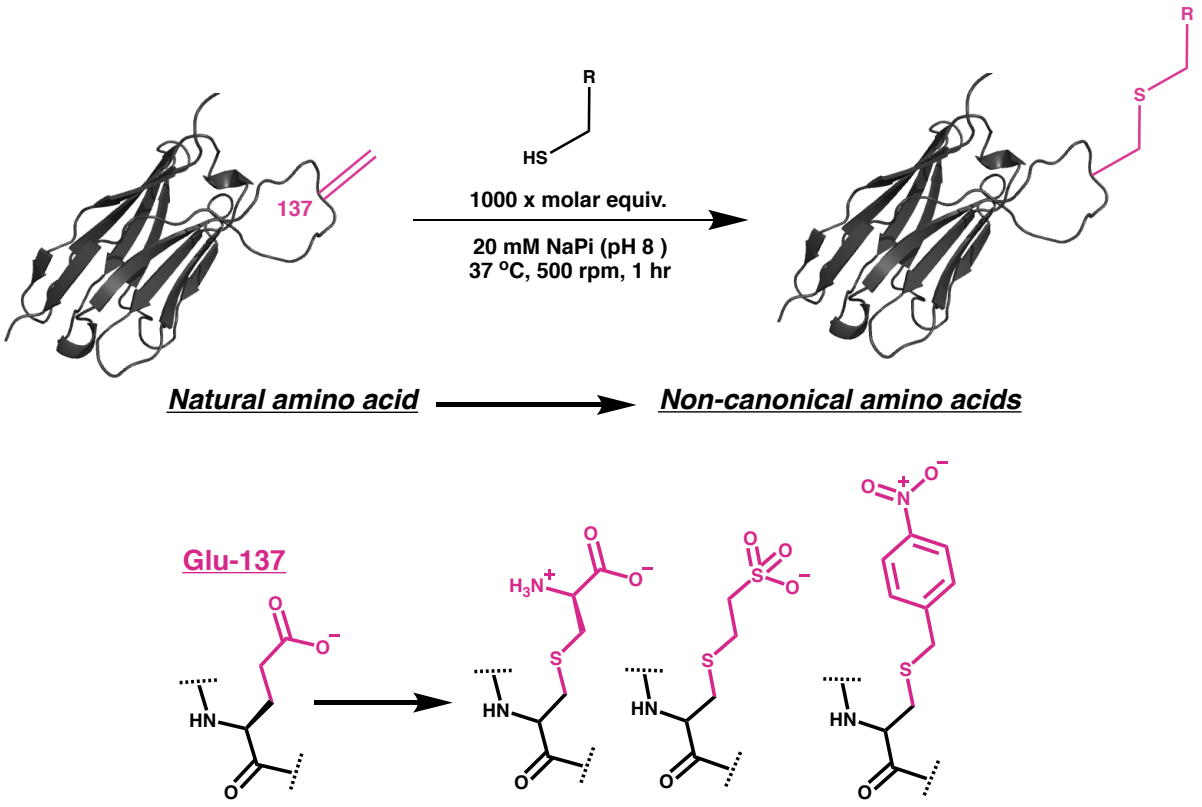


Figure 2.8. Aza-Michael additions with piperidine. (a) Initial reaction scheme for modifying DesAB-Aβ3-9 (T138Dha, K80E) with piperidine. (b) LC-MS spectrum of the homogenous product from the initial conditions. (c) LC-MS spectrum of the modified protein after dropping the molar equivalents of piperidine from 250 to 100 and extended the reaction time to 1 hr. (d) CD spectrums of the piperidine modified proteins.

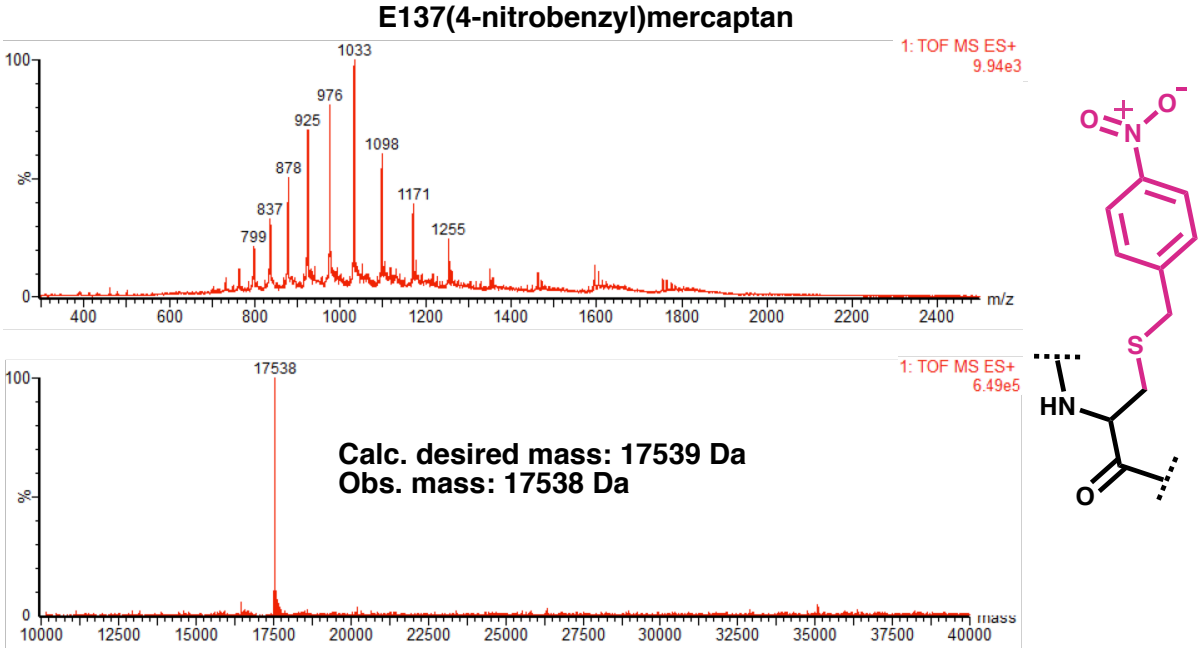
2.2.6. Side-chain exploration and activity screening at E137Dha

Dha at position 137 was one of the most reactive with β-mercaptoethanol (<15 min) and was at a position previously occupied by a negatively charged amino acid (glutamic acid). This led us to explore non-canonical side chains at this position containing negatively charged functional groups to see if modulating this property could lead to an activity increase. Three different chemical mutants were chosen to be screened: (4-nitrobenzyl)mercaptan, 2-mercaptoethanesulfonate (MESNA), and L-cysteine (**Figure 2.9a**). The conditions for conjugation were unchanged from the β-mercaptoethanol screen (20 mM NaPi [pH 8], 37 °C, 500 rpm) except that the reaction time was extended to an hour in case there was a variable reactivity amongst the attacking thiols. After 1 hr the LC-MS confirmed that all reactions had gone to completion with very homogeneous products (**Figure 2.9b-c**).

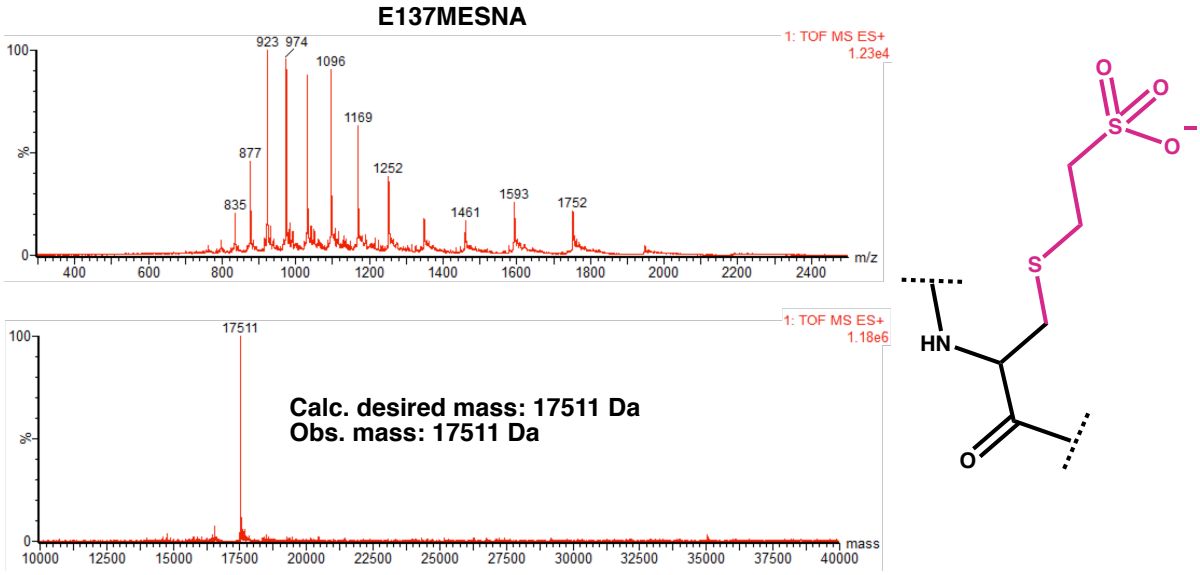
a



b


E137(4-nitrobenzyl)mercaptan
1: TOF MS ES+
9.94e3
1033
976
925
1098
878
1171
837
799
1255
m/z
1: TOF MS ES+
6.49e5
17538
Calc. desired mass: 17539 Da
Obs. mass: 17538 Da
mass

c


E137MESNA
1: TOF MS ES+
1.23e4
923
974
1096
1169
877
1252
835
1461
1593
1752
m/z
1: TOF MS ES+
1.18e6
17511
Calc. desired mass: 17511 Da
Obs. mass: 17511 Da
mass

d

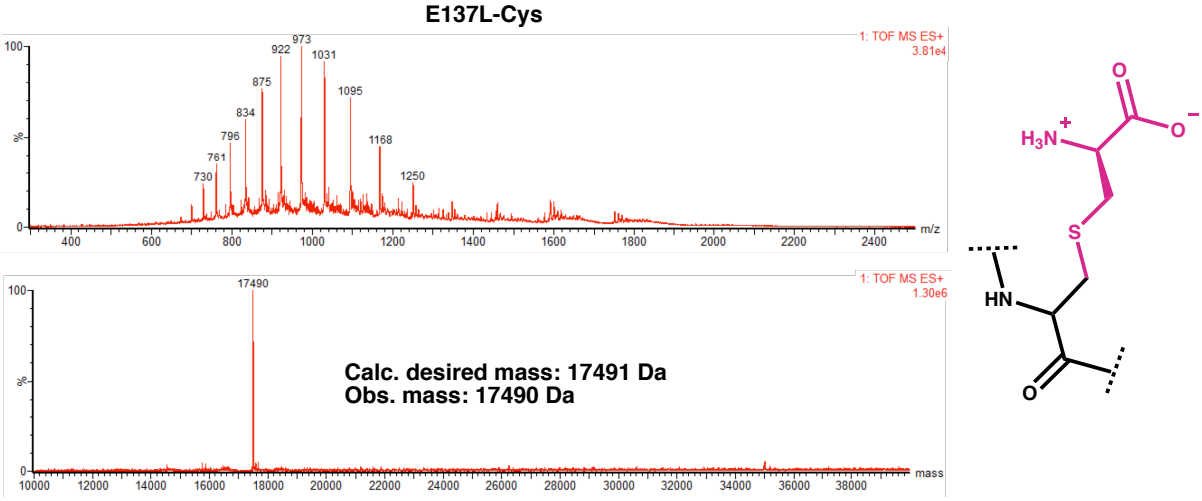


Figure 2.9. E137Dha reactions. (a) Reaction scheme for side-chain exploration at E137Dha through thia-Michael additions and the three chemical mutants generated. (b) LC-MS data for E137(4-nitrobenzyl)mercaptan confirming a homogeneous conversion. (c) LC-MS data for E137MESNA confirming a homogeneous conversion. (d) LC-MS data for E137L-Cys confirming a homogeneous conversion.

Before screening for activity, the structural integrity of the chemical mutants was assessed via CD (**Figure 2.10**). This confirmed that all of the mutants at position 137 displayed a high similarity to the starting scaffold.

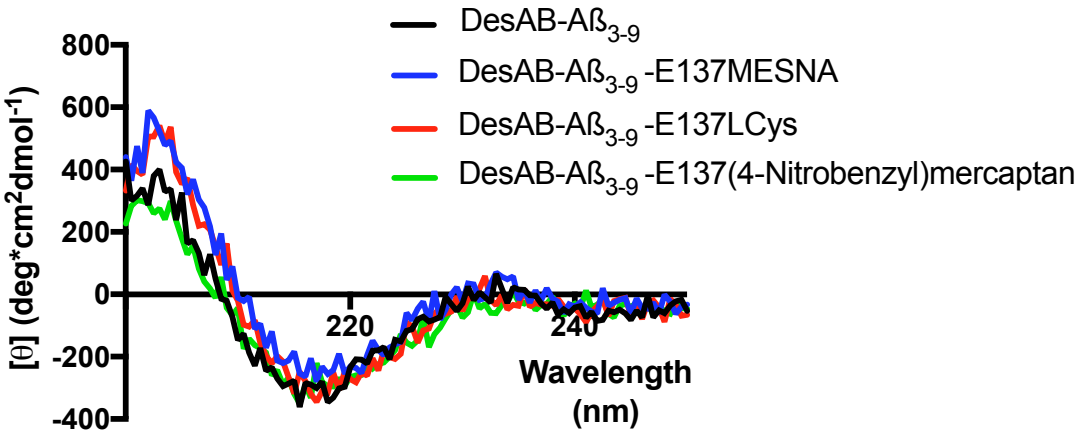
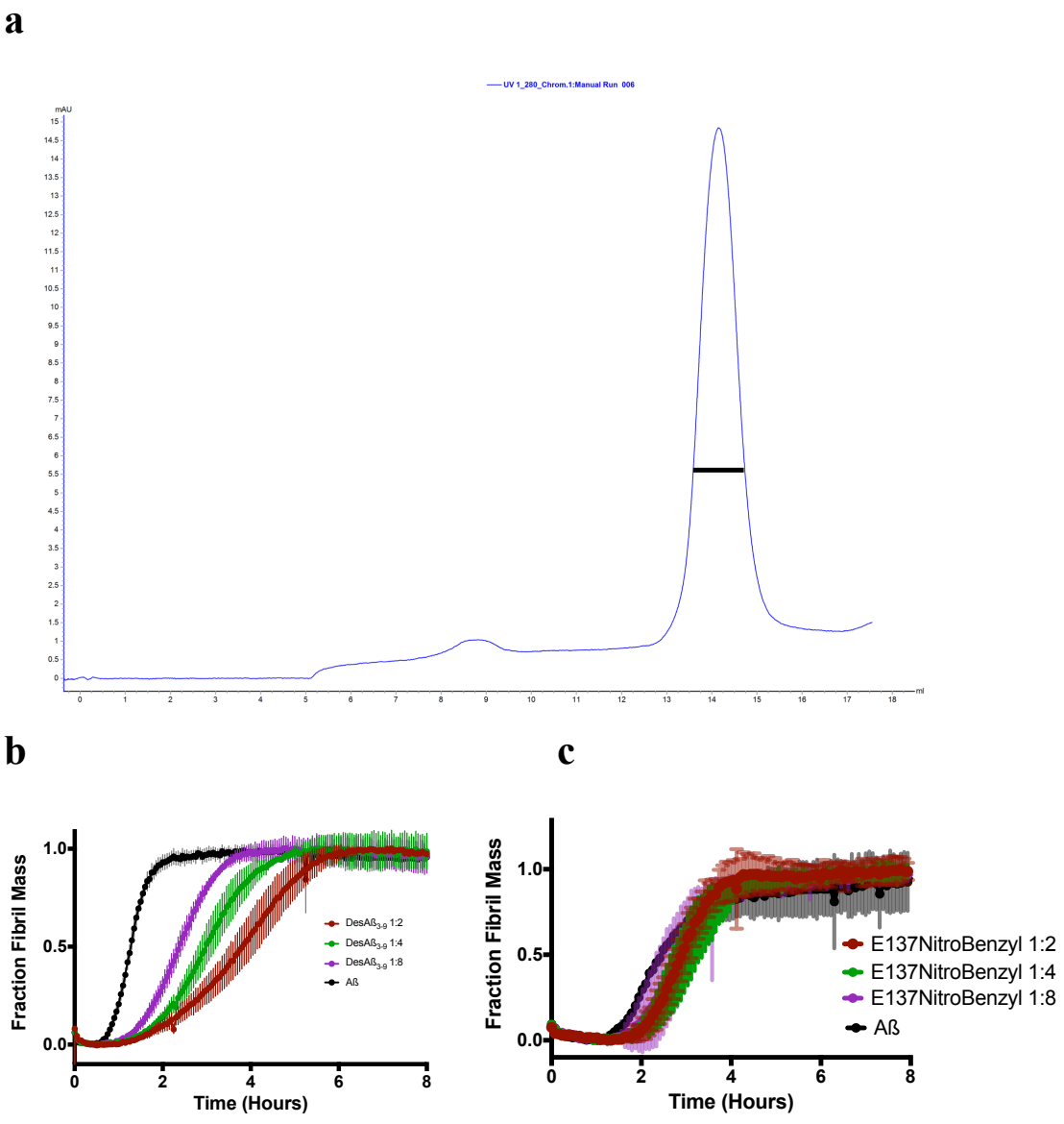


Figure 2.10. E137 chemical mutant's CD profiles. CD traces of DesAB-Aβ$_{3-9}$ and the various chemical mutants at position 137.

With the chemical mutants in hand and their integrity checked an activity screen was performed using an Aβ$_{42}$ anti-aggregation assay. To accurately assess the mechanism of inhibition, the ThT-based fluorescence of the aggregation of Aβ$_{42}$ under reference conditions is carried out alongside a serial dilution of the inhibitor in question. If the inhibitor is indeed

active and alters the aggregation profile, these perturbations can be analysed to elucidate the changes in the different rate constants caused by the inhibitor[83]. Lyophilised recombinant $A\beta_{42}$ graciously prepared by Swapan Preet and Ewa Klimont of the Centre for Misfolding Diseases. To prepare the $A\beta_{42}$ for the kinetic assays, a final SEC step was carried out after resuspending the lyophilised protein in 6 M guanidinium chloride in order to ensure the peptide was in its monomeric form at the beginning of the assay (**Figure 2.11a**). Using this approach, the chemical mutants were screened at three dilutions (1:2, 1:4 and 1:8 [antibody]:[$A\beta_{42}$]) in triplicate, with the concentration of $A\beta_{42}$ held constant at 1.5 μM. The modest activity of the starting DesAB-$A\beta_{3\text{-}9}$ can be seen for reference in **Figure 2.11b**. Unfortunately, all of the chemical mutants screened at position E137 ablated all anti-aggregation activity (**Figure 2.11c-e**). Therefore, no further modifications were pursued at this site.



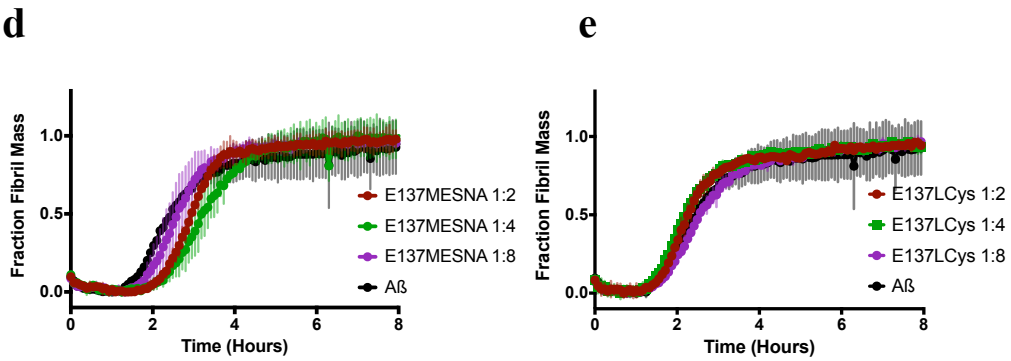
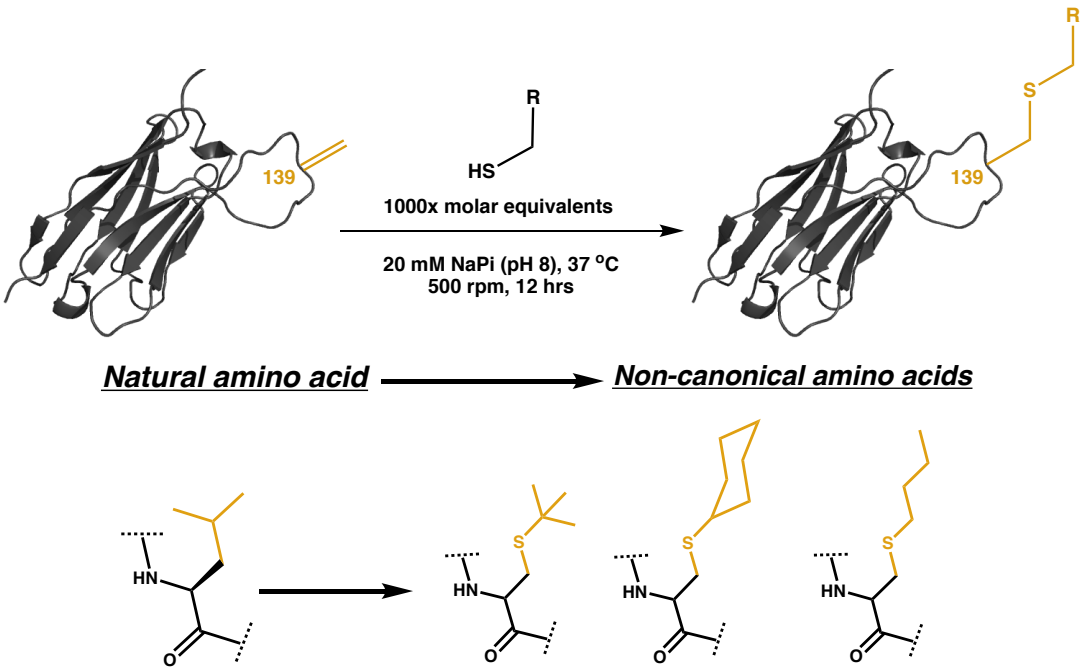


Figure 2.11. $A\beta_{42}$ purification and anti-aggregation screens for E137 chemical mutants. (a) Chromatogram at 280 nm for the final SEC of $A\beta_{42}$. The fraction collected for use in the aggregation assays is marked by the lack bar. (b) Aggregation curves for DesAB-$A\beta_{3\text{-}9}$. (c) Aggregation curves for DesAB-$A\beta_{3\text{-}9}$ (E137[4-nitrobenzyl]mercaptan). (d) Aggregation curves for DesAB-$A\beta_{3\text{-}9}$(E137MESNA). (e) Aggregation curves for DesAB-$A\beta_{3\text{-}9}$(E137L-Cys). The Aβ concentration for each reaction is 1.5 μM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

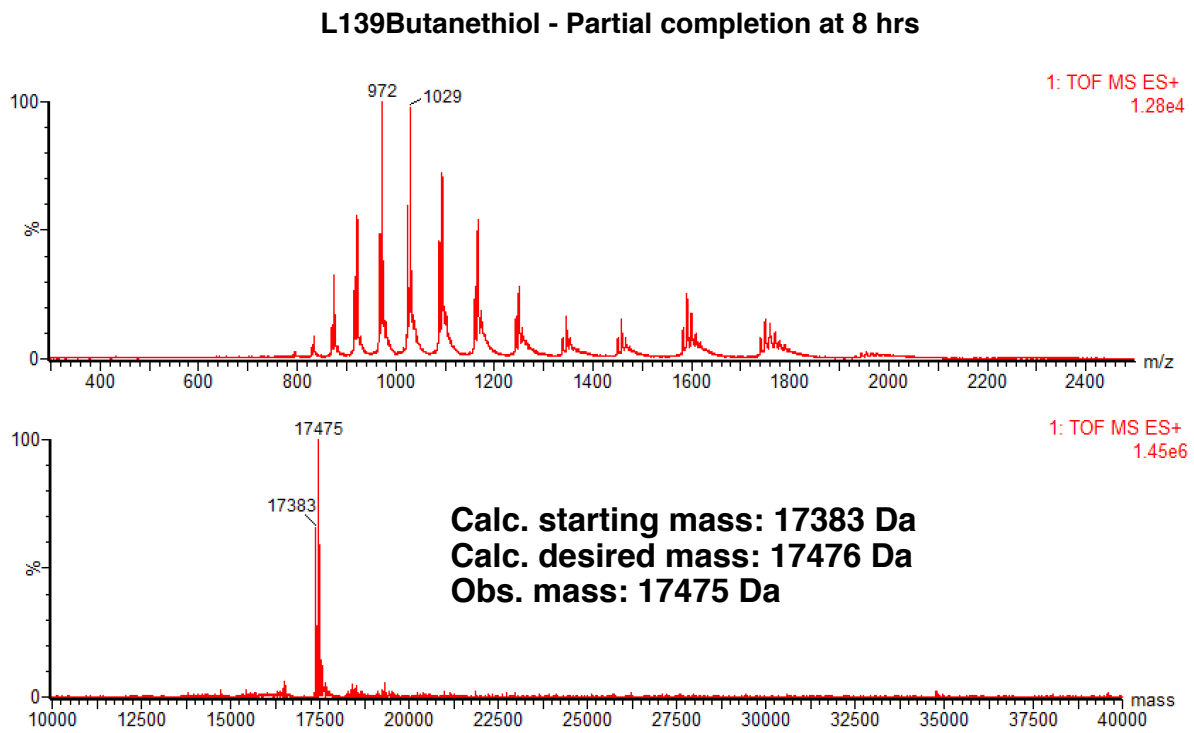
2.2.7. Side-chain exploration and activity screening at L139Dha

Position 139 was equally reactive as 137 to the modification of Dha by β-mercaptoethanol and was originally occupied by leucine. Therefore, we chose to explore the installation of a variety of non-canonical aliphatic sidechains at this position. The three initial aliphatic molecules pursued for conjugation were: tert-butanethiol, hexanethiol, and 1-butanethiol (**Figure 2.12a**). These chemicals produce very potent noxious odours and great care must be taken when handling them outside of a fume hood. While β-mercaptoethanol was capable of fully conjugating to L138Dha in less than 15 mins, the aliphatic thiols took considerably longer to reach completion. After reacting for 8 hrs, all three chemical mutants had only achieved partial conversion to the desired product (**Figure 2.12b**). To achieve full conversions the reaction time had to be extended to 12 hrs (**Figure 2.12c-e**). The discrepancy in the reactivity of the aliphatic thiols compared to β-mercaptoethanol could arise from the different miscibility profiles of the thiols. The volatility of the aliphatic thiols also likely meant that less than 1000x molar equivalents was added to the reaction. If these chemical mutations were to be attempted on a less thermally stable protein, the addition of an organic co-solvent (such as *N,N*-dimethylformamide [DMF]) may speed up the reaction.

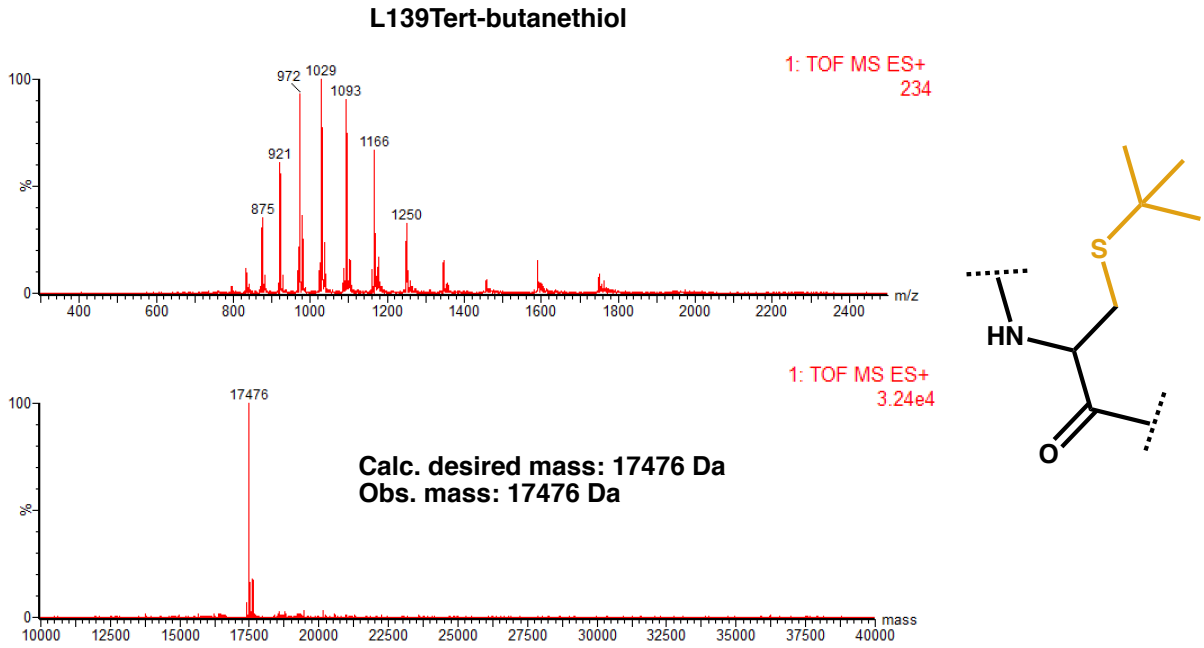
a


139
R
HS
1000x molar equivalents
20 mM NaPi (pH 8), 37 °C
500 rpm, 12 hrs
R
S
139
Natural amino acid
Non-canonical amino acids
HN
O
S

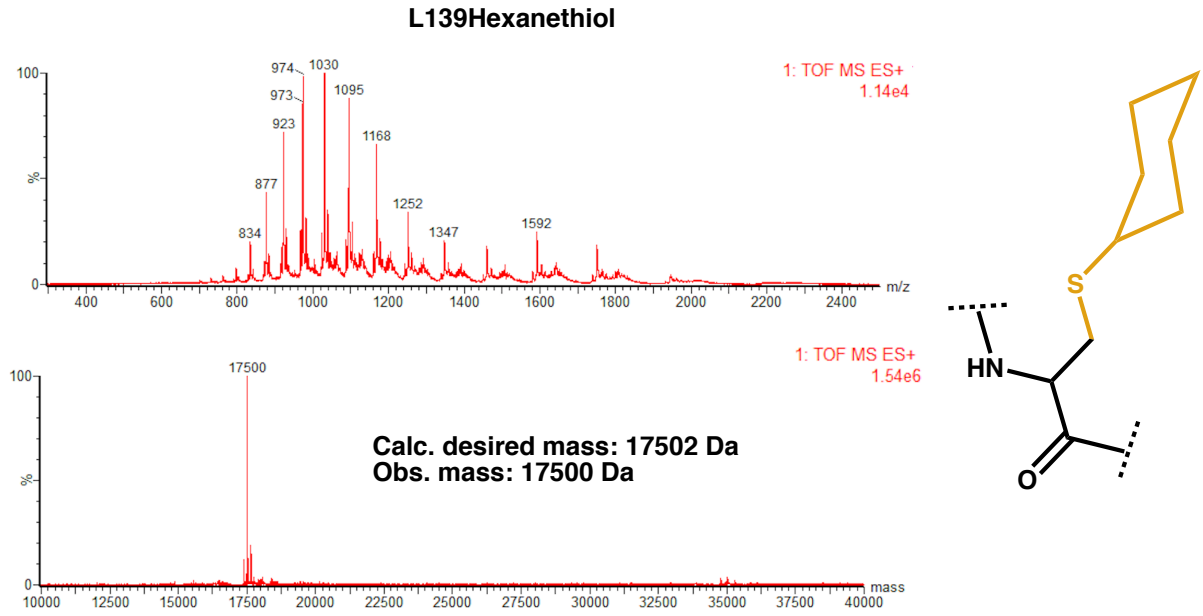
b


L139Butanethiol - Partial completion at 8 hrs
1: TOF MS ES+
1.28e4
100
%
0
972
1029
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
1.45e6
17475
17383
Calc. starting mass: 17383 Da
Calc. desired mass: 17476 Da
Obs. mass: 17475 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass

c


L139Tert-butanethiol
1: TOF MS ES+
234
100
%
0
972
1029
1093
921
1166
875
1250
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
3.24e4
17476
Calc. desired mass: 17476 Da
Obs. mass: 17476 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass
S
HN
O

d


L139Hexanethiol
1: TOF MS ES+
1.14e4
100
%
0
974
1030
973
1095
923
1168
877
1252
834
1347
1592
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
1.54e6
17500
Calc. desired mass: 17502 Da
Obs. mass: 17500 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass
S
HN
O

e

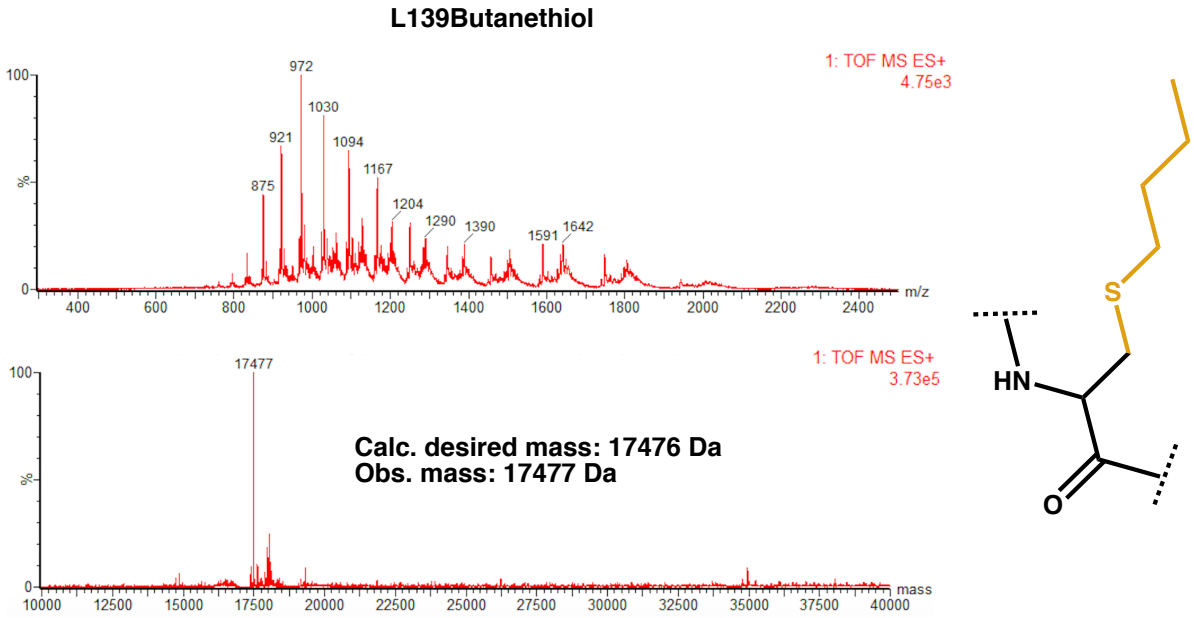


Figure 2.12. L139Dha reactions. (a) Reaction scheme for side-chain exploration at L139Dha through thia-Michael additions and the three chemical mutants generated. (b) Representative timepoint after reacting for 8 hrs for the conjugation of 1-butanethiol to L139Dha, the other two aliphatic thiols had comparable conversions at this time. (c) LC-MS data for L139Tert-butanethiol confirming a homogeneous conversion. (d) LC-MS data for L139Hexanethiol confirming a homogeneous conversion. (e) LC-MS data for L139Butanethiol confirming a homogeneous conversion.

Just as with the E137 chemical mutants, the aliphatic L139 mutants had their structural integrity assessed by CD. This confirmed that even after the 12-hr reaction time, the chemical mutants were still folded correctly and could then be screened for their anti-aggregation activity (**Figure 2.13**).

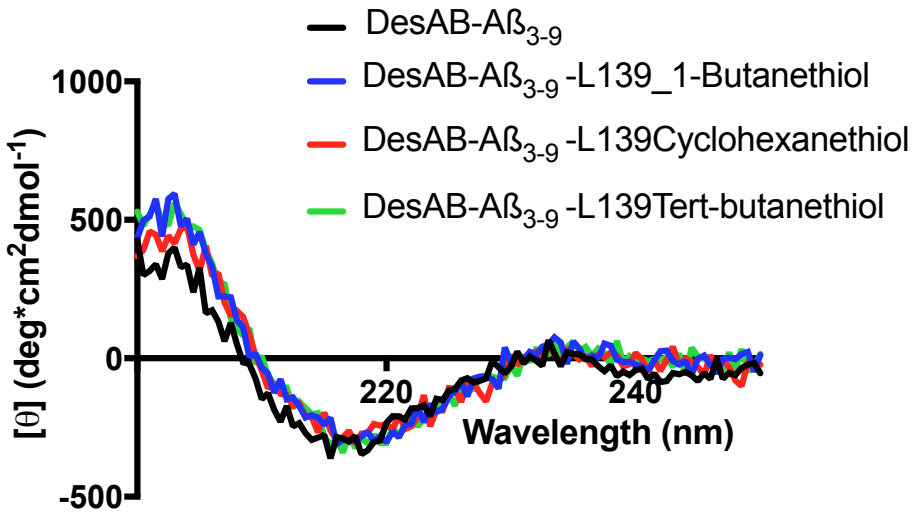


Figure 2.13. L139 chemical mutant's CD profiles. CD traces of DesAB-Aβ$_{3-9}$ and the various chemical mutants at position 139.

These three chemical mutants had their anti-aggregation activity assessed in the same manner as the E137 mutants. Disappointingly, none of these larger aliphatic groups had any significant benefit for the anti-aggregation activity (**Figure 2.14a-c).** Therefore, no further mutants were pursued at this position.

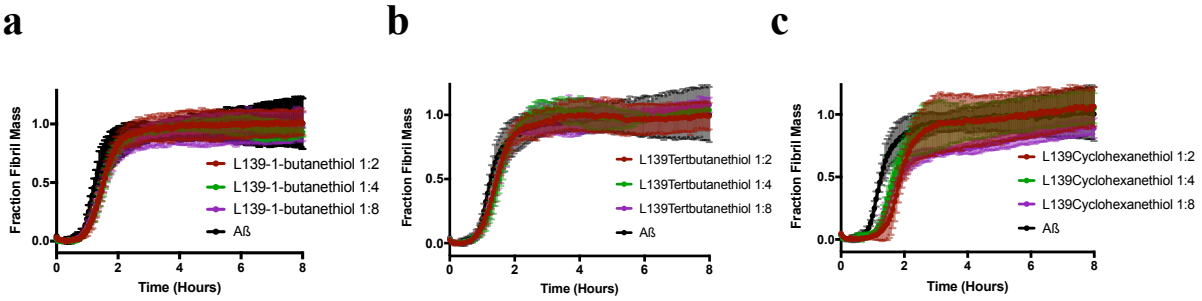


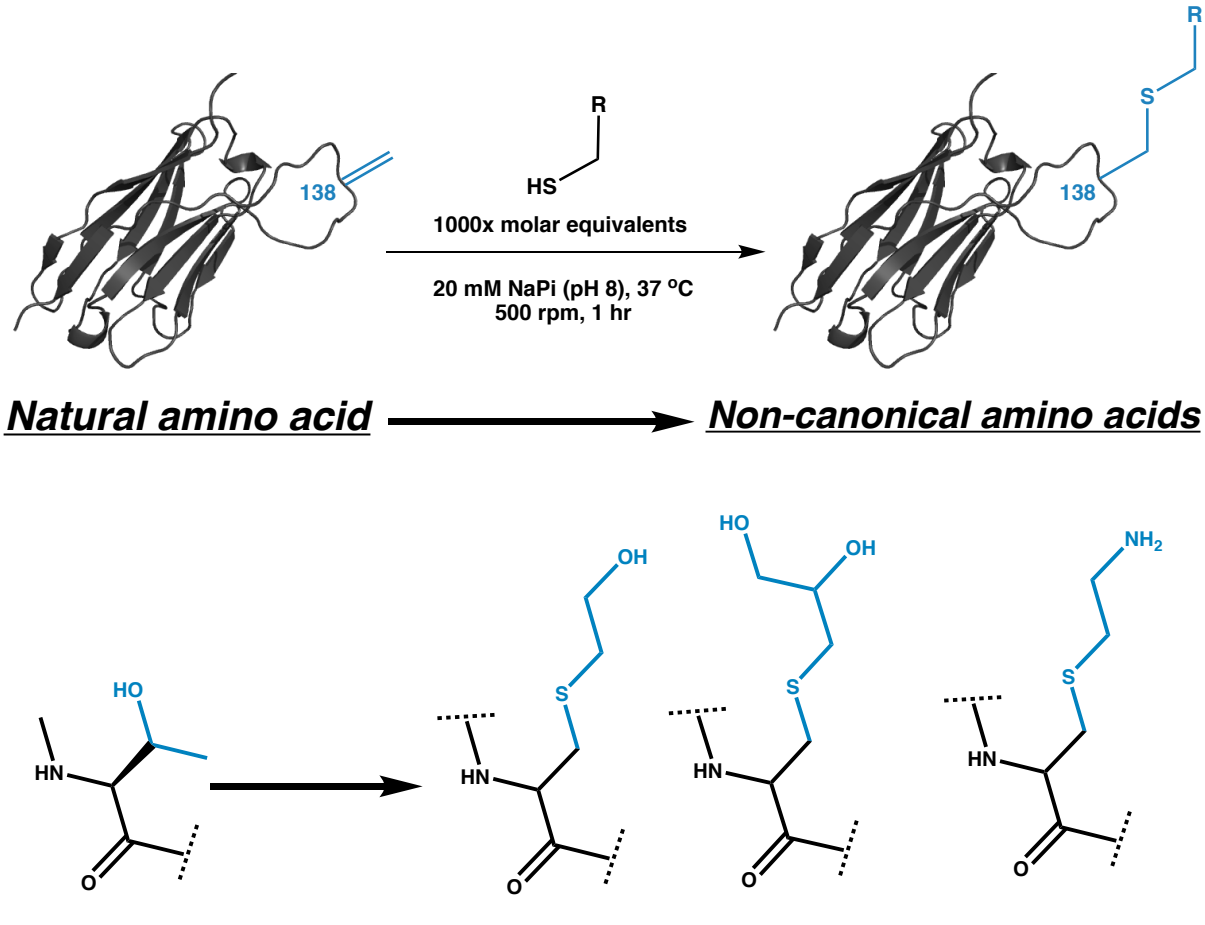
Figure 2.14. Aβ$_{42}$ anti-aggregation screens for L139 chemical mutants. (a) Aggregation curves for DesAB-Aβ$_{3-9}$ (L139-1-butanethiol). (b) Aggregation curves for DesAB-Aβ$_{3-9}$(L139Tert-butanethiol). (c) Aggregation curves for DesAB-Aβ$_{3-9}$(L139Cyclohexanethiol). The Aβ concentration for each reaction is 1.5 μM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

2.2.8. Initial side-chain exploration and activity screening at T138Dha

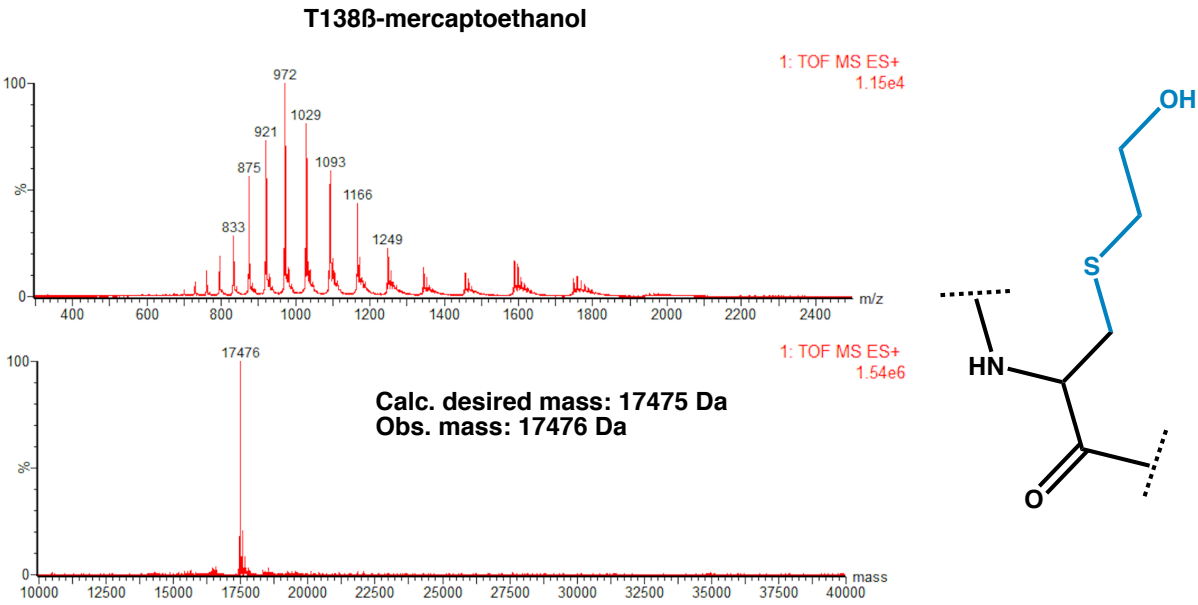
Dha at position T138 was the least reactive with β-mercaptoethanol compared to the other two positions chosen for sidechain exploration. With threonine being the original sidechain in the designed sequence, we decided to explore extending the polar group, adding

more polar groups, and altering the heteroatom. To achieve these alterations, we conjugated β-mercaptoethanol, 1-thioglycerol, and cysteamine to T138Dha. Since the other two thiols were very similar in properties to β-mercaptoethanol, the same reaction conditions and time that was initially determined in the reactivity screen was used for each reaction (**Figure 2.15a**). All reactions yielded homogenous desired products by 1 hr (**Figure 2.15b-d**).

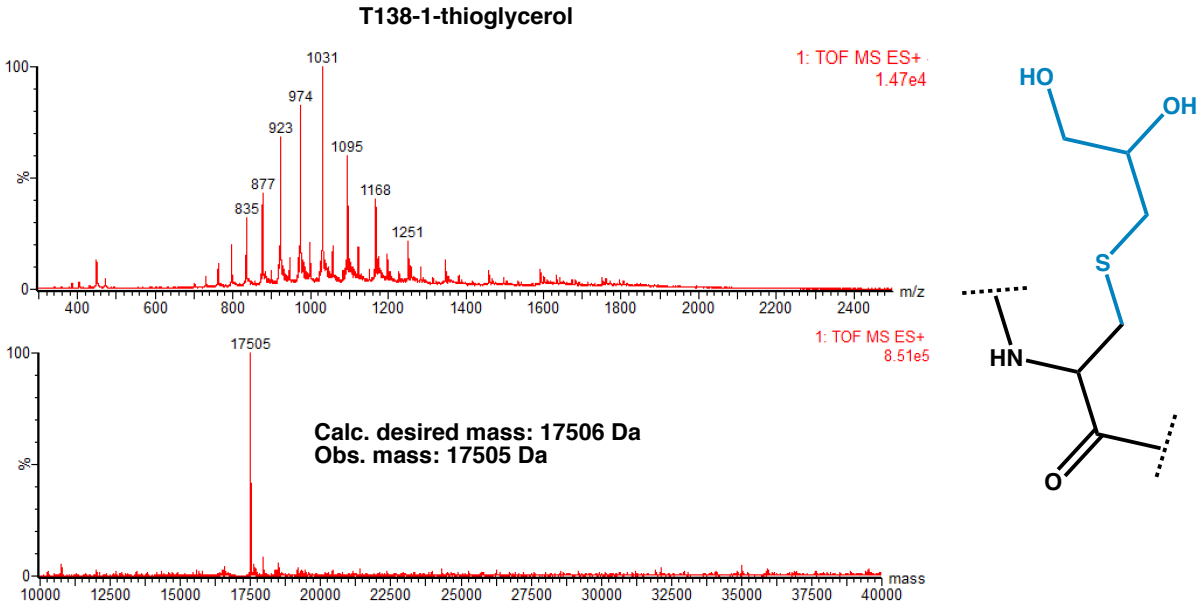
a



b


T138ß-mercaptoethanol
1: TOF MS ES+
1.15e4
100
%
0
972
1029
921
875
1093
1166
833
1249
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
1.54e6
17476
Calc. desired mass: 17475 Da
Obs. mass: 17476 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass
OH
S
HN
O

c


T138-1-thioglycerol
1: TOF MS ES+
1.47e4
100
%
0
1031
974
923
1095
877
1168
835
1251
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z
1: TOF MS ES+
8.51e5
17505
Calc. desired mass: 17506 Da
Obs. mass: 17505 Da
10000
12500
15000
17500
20000
22500
25000
27500
30000
32500
35000
37500
40000
mass
HO
OH
S
HN
O

d

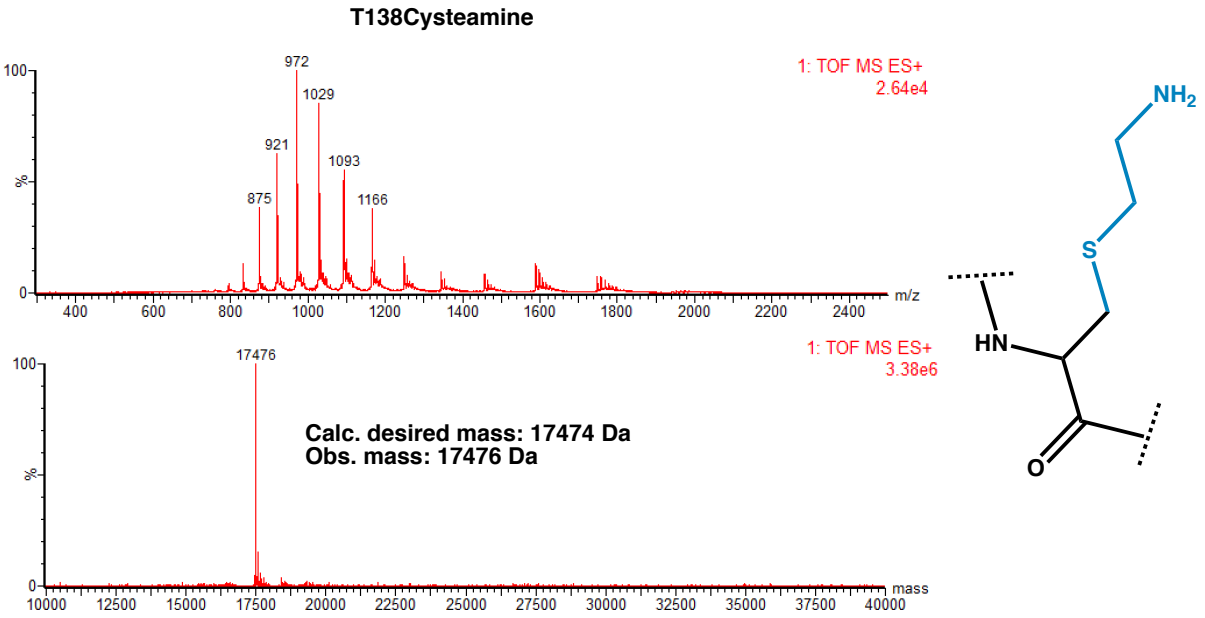


Figure 2.15. T138Dha reactions. (a) Reaction scheme for initial side-chain exploration at T138Dha through thia-Michael additions and the three chemical mutants generated. (b) LC-MS data for T138β-mercaptoethanol confirming a homogeneous conversion. (d) LC-MS data for T138-1-thioglycerol confirming a homogeneous conversion. (e) LC-MS data for T138Cysteamine confirming a homogeneous conversion.

Just as with the chemical mutants at E137 and L139, the T138 mutants had their structural integrity confirmed by CD. Each mutant had a CD profile comparable to that of DesAB-Aβ$_{3-9}$, and thusly they could all be screened for their anti-aggregation activity (**Figure 2.16**).

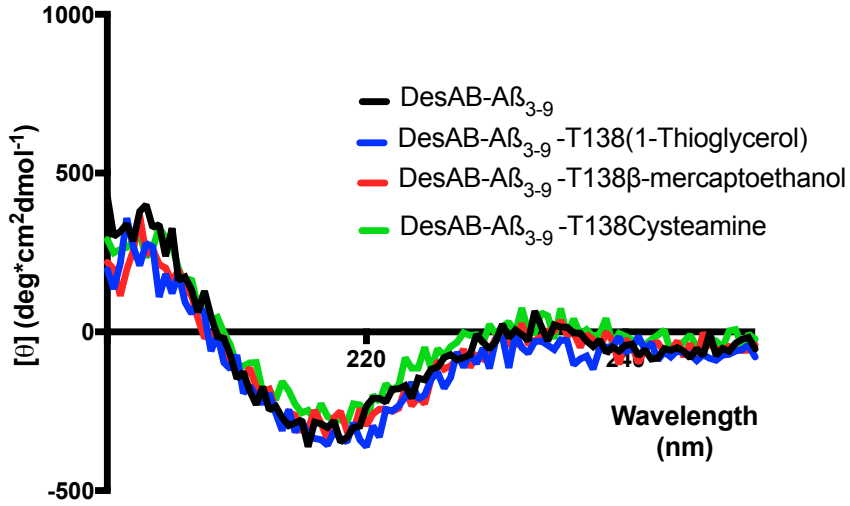


Figure 2.16. T138 chemical mutant's CD profiles. CD traces of DesAB-Aβ$_{3-9}$ and the various chemical mutants at position T138.

The T138 chemical mutants had their anti-aggregation activity screened in the same manner as the previous chemical mutants. This revealed that while the β-mercaptoethanol and 1-thioglycerol chemical mutants ablated activity, the cysteamine mutant displayed a dose-dependent alteration of the aggregation profile (**Figure 2.17a-d**).

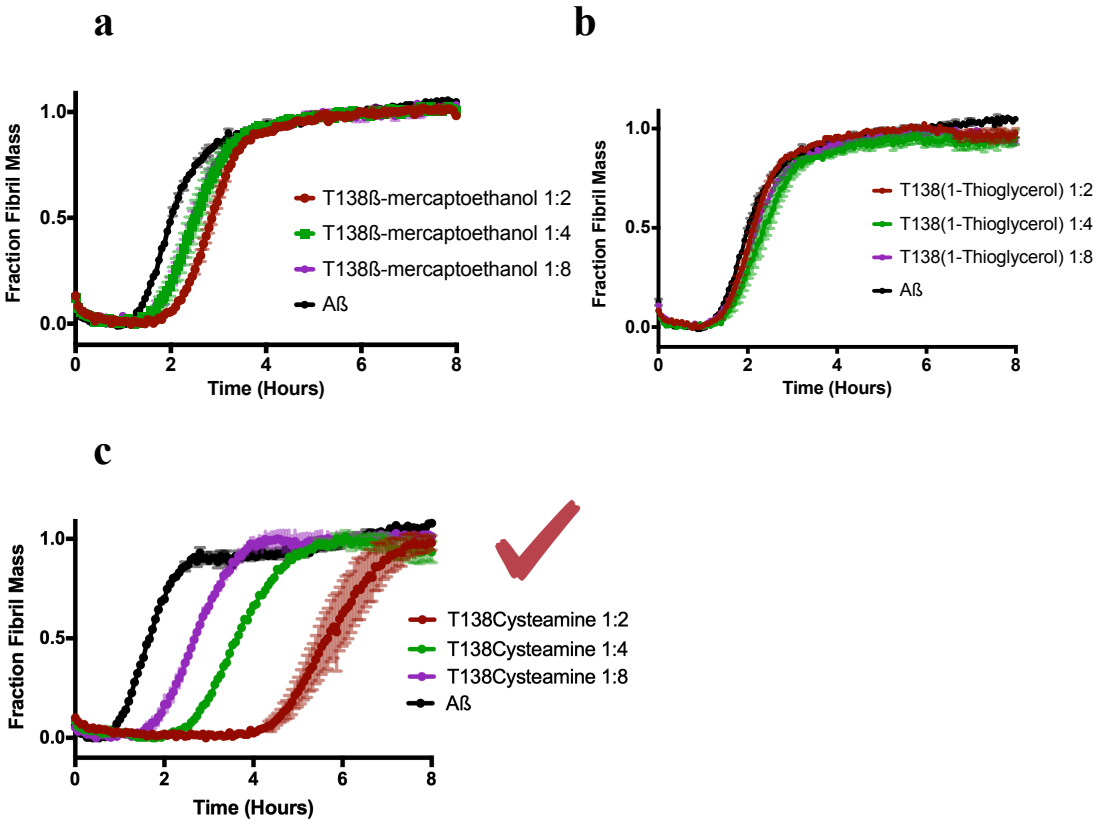


Figure 2.17. $A\beta_{42}$ anti-aggregation screens for T138 chemical mutants. (a) Aggregation curves for DesAB-$A\beta_{3-9}$ (T138β-mercaptoethanol). (b) Aggregation curves for DesAB-$A\beta_{3-9}$ (T138-1-thioglycerol). (c) Aggregation curves for DesAB-$A\beta_{3-9}$ (T138Cysteamine). The Aβ concentration for each reaction is 1.5 μM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

To confirm that the cysteamine mutation at T138 did indeed enhance the anti-aggregation activity, we sought to elucidate the effect of the mutation on the microscopic steps of the aggregation process. Using the integrated kinetic rate laws described in **Section 1.1.5**, we determined the antibody's effect on the global parameters k_+k_n and k_+k_2 (where again k_+ is the elongation rate, k_n is the nucleation rate, and k_2 is the secondary nucleation rate). Each curve within the dilution series of both DesAB-$A\beta_{3-9}$ and DesAB-$A\beta_{3-9}$(T138Cysteamine) was fit

well by the model (**Figure 2.18a & c**) and revealed that the cysteamine mutation does indeed enhance the inhibition of the primary nucleation reaction (**Figure 2.18b & d**).

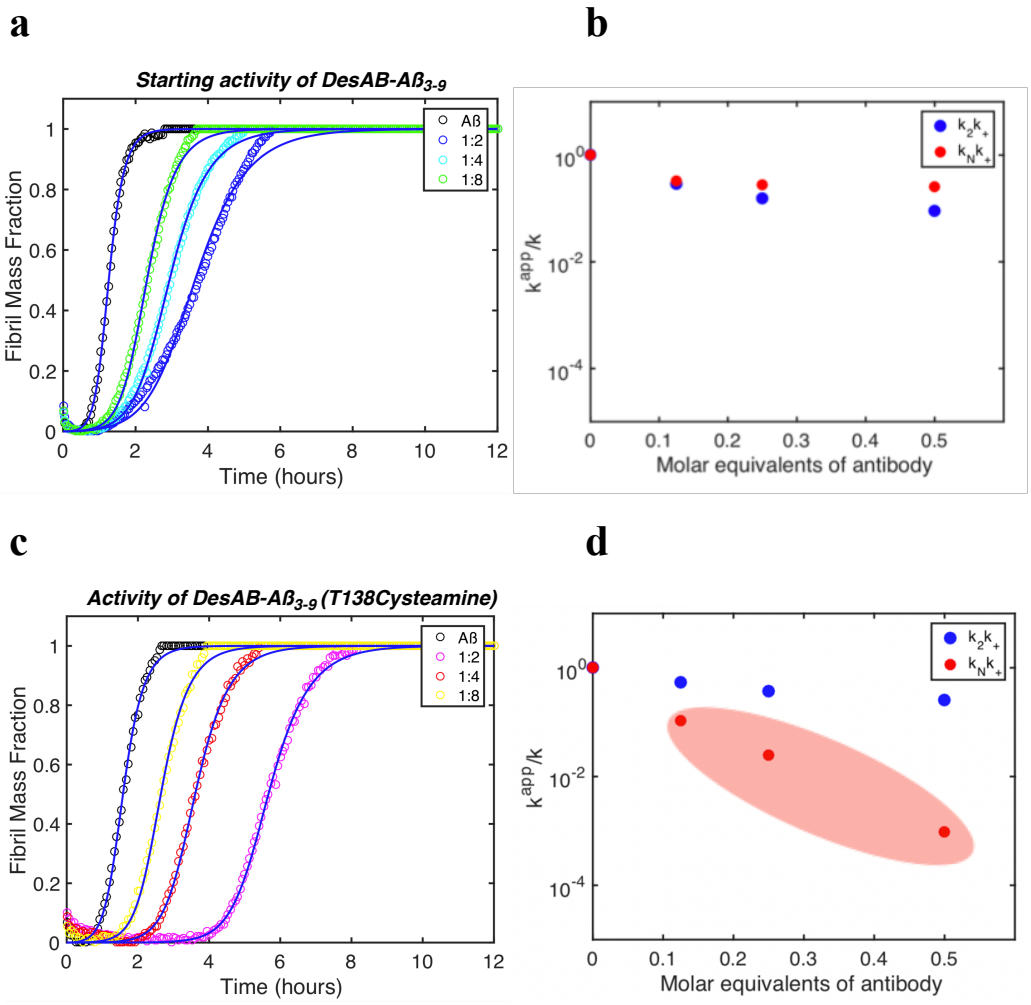
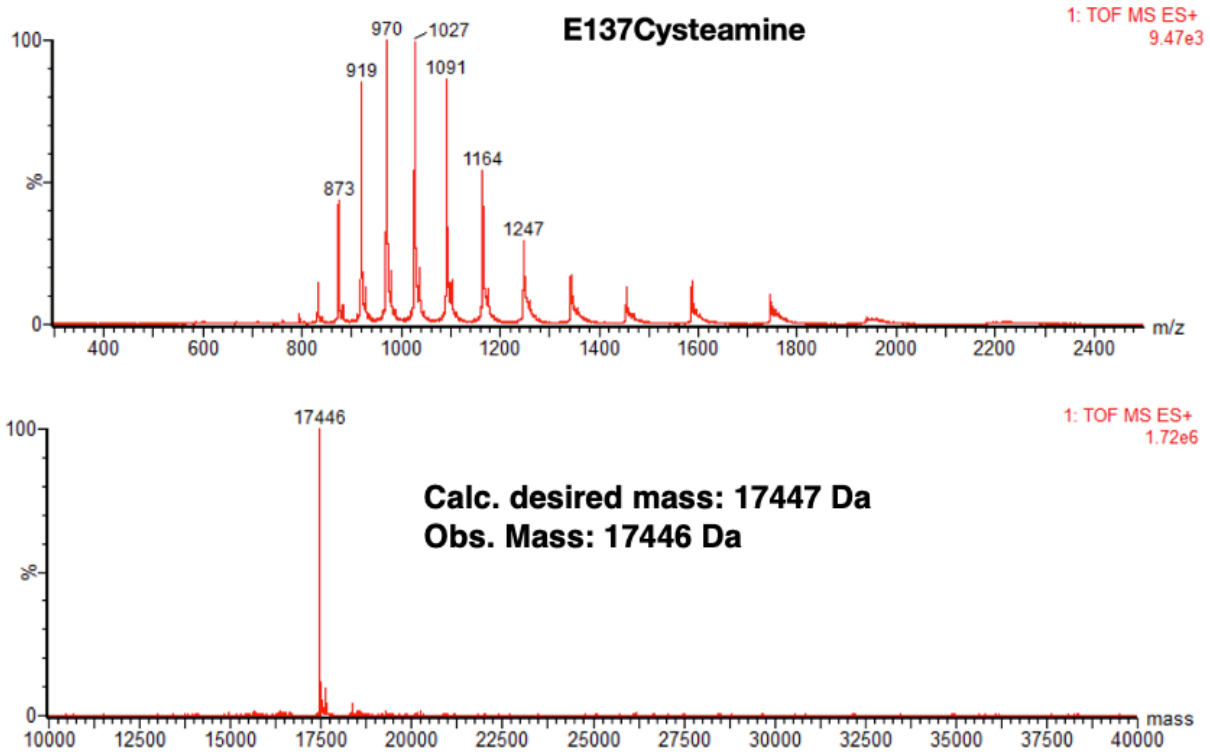


Figure 2.18. DesAB-Aβ$_{3-9}$(T138Cysteamine) has enhanced inhibition of primary nucleation. (a) Normalised aggregation curves for the dilution series of DesAB-Aβ$_{3-9}$ (open coloured circles) and fits of the data using the integrated rate laws (solid blue lines). (b) Plot of the individual global aggregation parameters for each of the DesAB-Aβ$_{3-9}$ dilutions relative to the Aβ alone parameters. (c) Normalised aggregation curves for the dilution series of DesAB-Aβ$_{3-9}$ (T138Cysteamine) (open coloured circles) and fits of the data using the integrated rate laws (solid blue lines). (d) Plot of the individual global aggregation parameters for each of the DesAB-Aβ$_{3-9}$(T138Cysteamine) dilutions relative to the Aβ alone parameters, red shaded circle highlights the potentiated reduction of the k_+k_n parameter.

After confirming the activity enhancement with the kinetic model, a few more exploratory controls had to be conducted. To investigate if this potentiation by cysteamine is dependent on the site of installation we created chemical mutants with it conjugated to both E137 and L139 positions using the same reaction conditions as for T138. Cysteamine was efficiently installed at both sites and created homogenous products (**Figure 2.19a & b**). Screening these two mutants for their aggregation activity revealed that neither recapitulate the

activity of the mutation at position T138, confirming that the potentiation by the cysteamine sidechain is position dependent (**Figure 2.19c & d**).

a



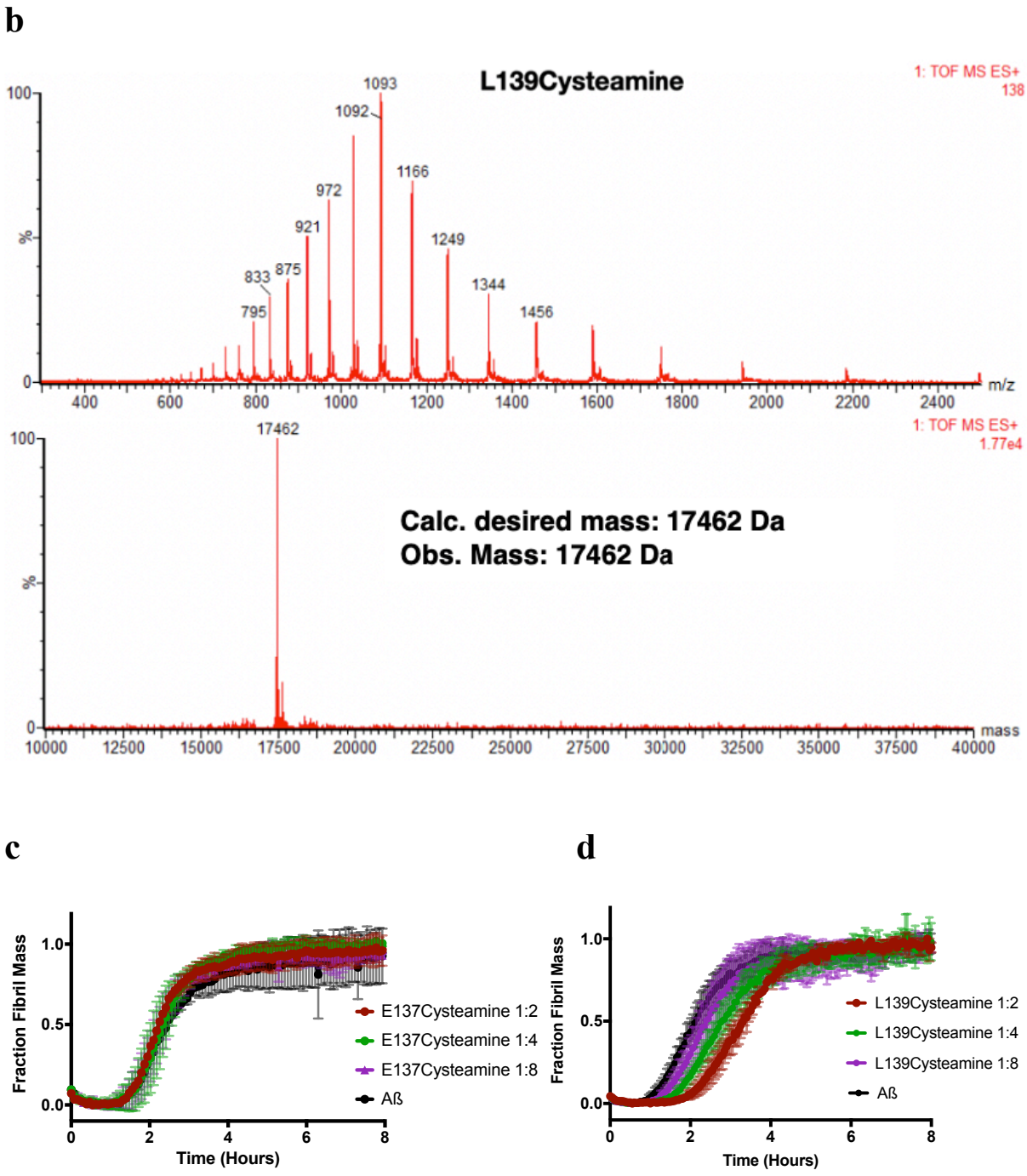
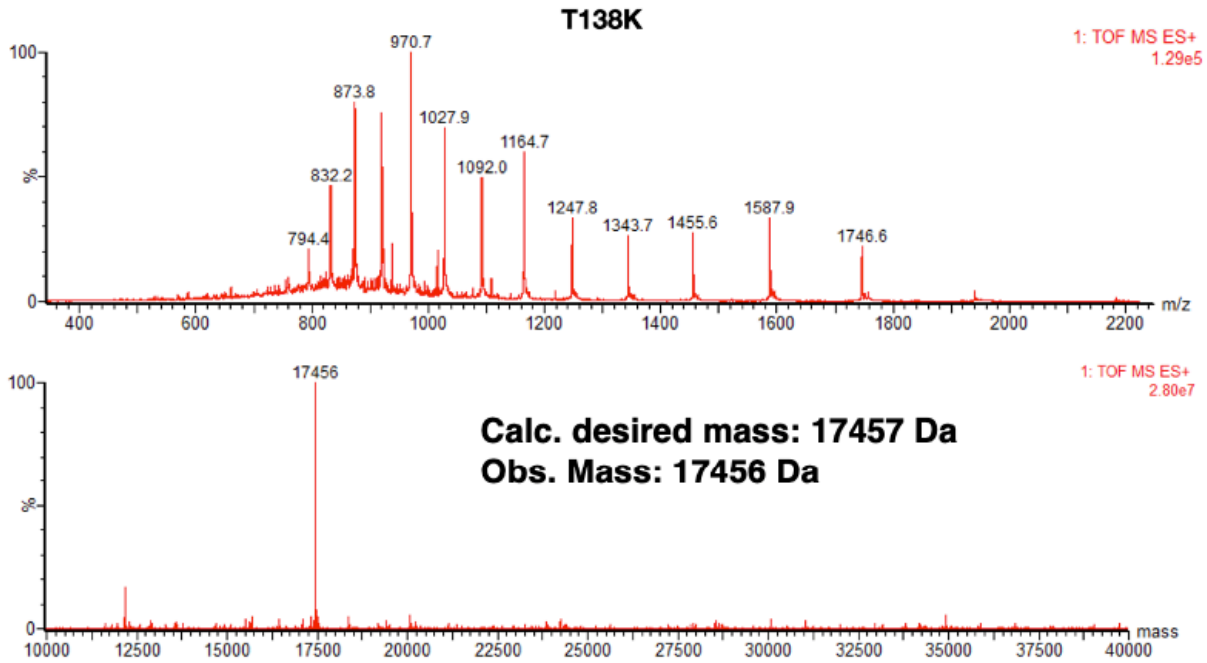


Figure 2.19. Creation and investigation cysteamine chemical mutants at positions E137 and L139. (a, b) LC-MS spectrums confirming the clean production of both DesAB-Aβ$_{3-9}$ (E137Cysteamine) and (L139Cysteamine) chemical mutants, respectively. (c, d) Anti-aggregation screens for cysteamine mutants at both positions, confirming neither recapitulate the activity observed at position T138. The Aβ concentration for each reaction is 1.5 µM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

Of the canonical amino acids, the cysteamine chemical mutant is most similar to lysine. The only differences are the thioether linkage at the γ position and the creation of a mixture of

epimers that is inherent when using Dha (see section **1.3.4**). With this high degree of similarity, we were curious as to if the chemical installation of cysteamine was necessary to enhance DesAB-Aβ_{3-9}'s activity or if a simple genetic mutation to lysine at position 138 would also potentiate activity. DesAB-Aβ_{3-9}(T138K) was thusly created using conventional site directed mutagenesis and purified in the same manner as previously described, yielding the expected product (**Figure 2.20a**). To account for any changes that may have occurred during the reaction process, the T138K mutant was subjected to the same prolonged incubations at 37 ºC in the same buffers that were used to generate T138Cysteamine. At the end of this, the CD profile of T138K was very similar to both DesAB-Aβ_{3-9} and T138Cysteamine (**Figure 2.20b**). However, the anti-aggregation screen with T138K revealed that it had almost no activity (**Figure 2.20c**). This finding could be a result of the fact that addition at Dha creates an epimeric mixture of D and L sidechains, or possibly from some subtle conformational changes the CDR3 undergoes during the transformations. Regardless of the reasons, in summary cysteamine proved to be the first potentiating chemical mutant in the initial screen at the three positions of the CDR3 loop. Its enhancing effect is position dependent and cannot be recapitulated by the most similar canonical amino acid. Serendipitously, the terminal amine of cysteamine provides an excellent platform to explore further derivatives and try to further improve the anti-aggregation activity.

a



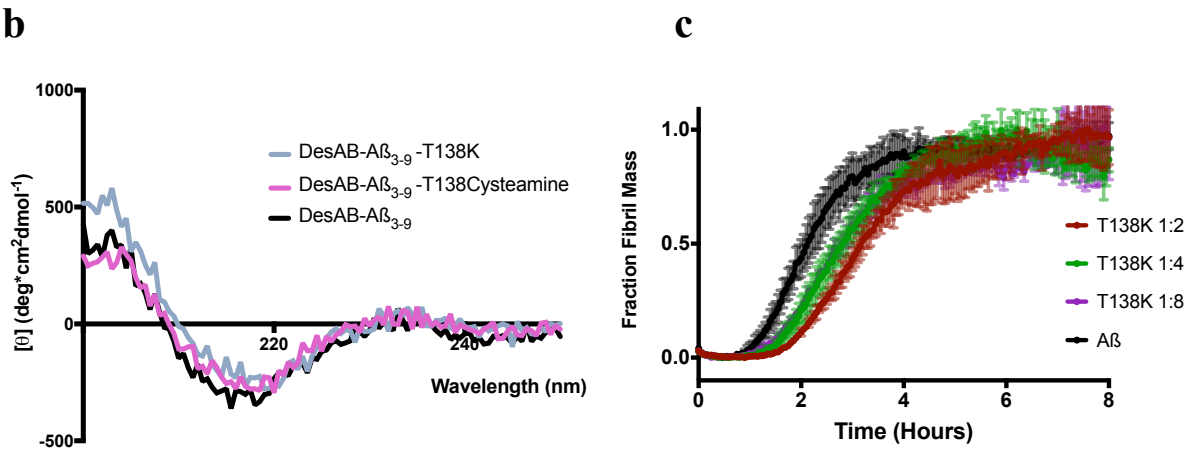
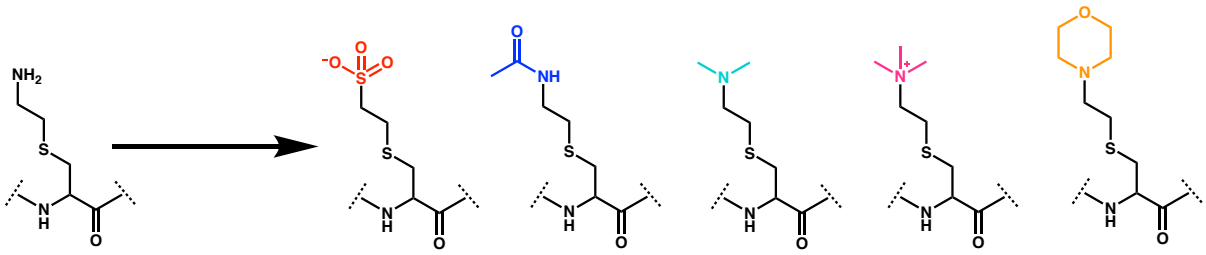


Figure 2.20. Creation and characterisation of DesAB-Aβ$_{3-9}$(T138K). (a) LC-MS confirming the desired mass. (b) CD spectrums of T138K and T138Cysteamine compared to the starting antibody's profile, verifying they have similar structures. (c) Aβ anti-aggregation screen for T138K. The Aβ concentration for each reaction is 1.5 μM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

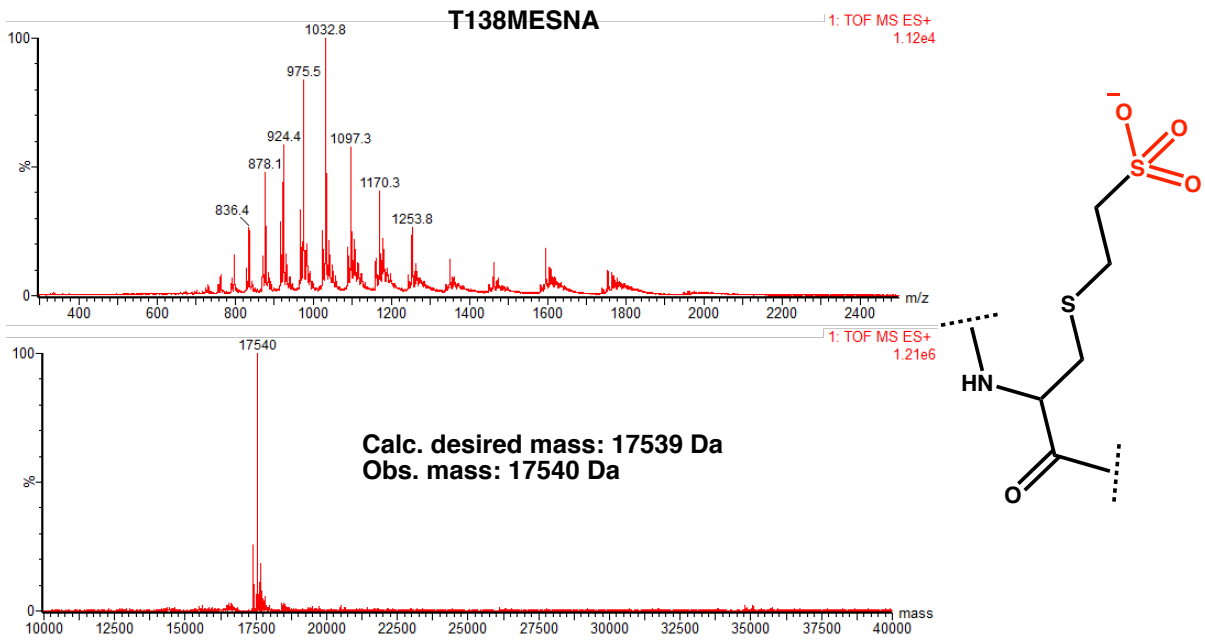
2.2.9. Derivatisation of T138Cysteamine

Small molecule SAR studies typically pursue multiple rounds of derivatisation before the desired biological activity is achieved. The terminal amine of our cysteamine hit provides an ideal point for the further exploration of subtle chemical alterations to further potentiate the anti-aggregation activity of DesAB-Aβ$_{3-9}$. Based on what derivatives were commercially available or could be easily synthesized, we created a panel of chemical mutants that had a range of physiochemical changes based on the cysteamine scaffold (**Figure 2.21a**). These changes included altering the charge of the side chain, such as removing it through the addition of an acetyl group, inverting it with a sulphate group, and fixing it through tri-methylation. Other changes included dimethylation and the addition of a larger heterocycle. The only desired derivative that was not commercially available was *N,N,N*-trimethylcysteamine, which had to by synthesised according to an established protocol (see **Materials and Methods**). All of these derivatives achieved homogenous conjugation to T138Dha under the same conditions used for cysteamine. (**Figure 2.21b-f**).

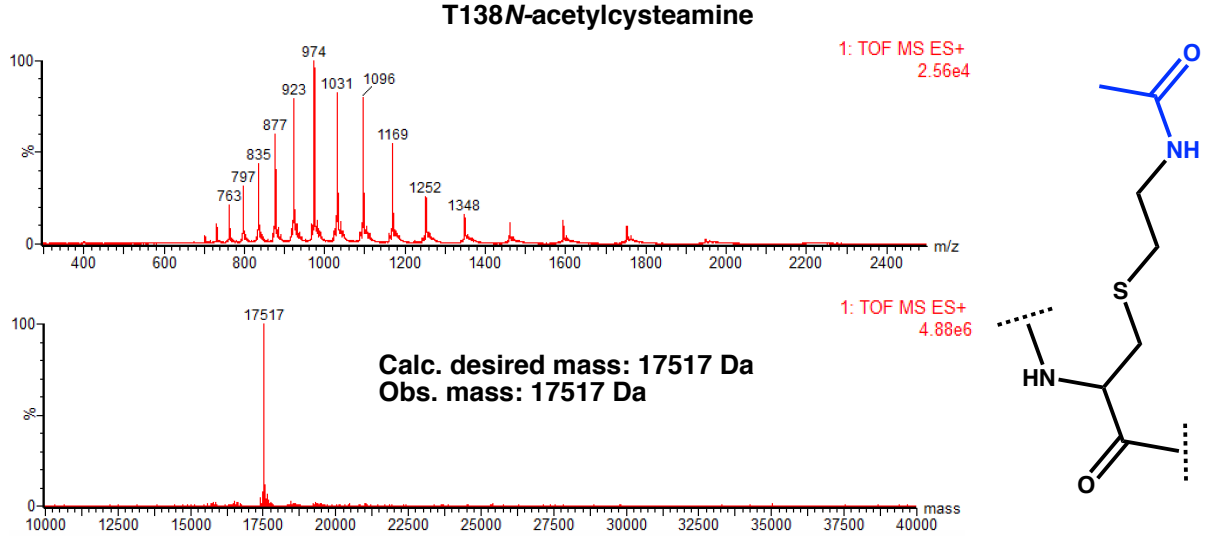
a



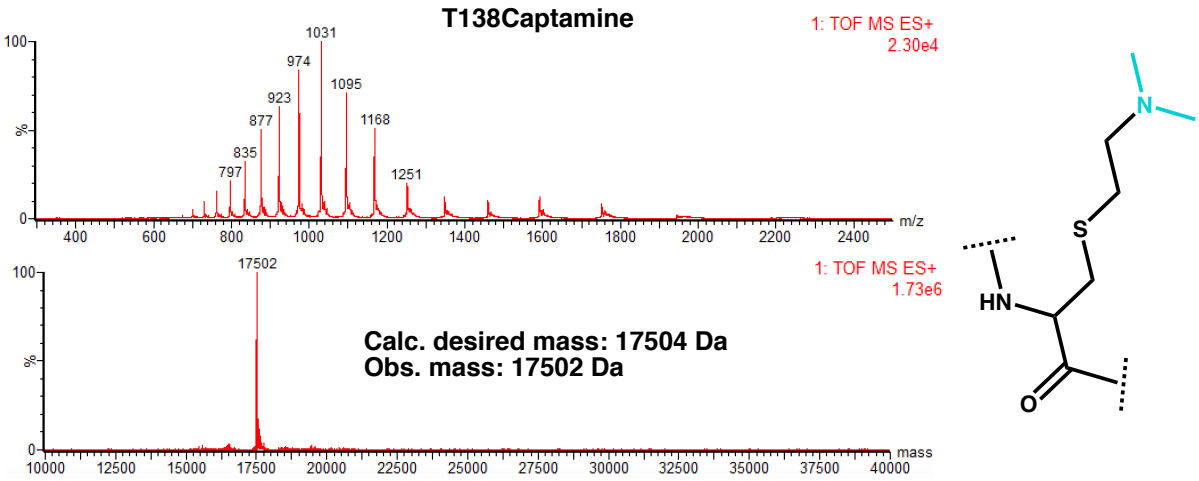
b


T138MESNA
1: TOF MS ES+
1.12e4
1032.8
975.5
924.4
878.1
836.4
1097.3
1170.3
1253.8
m/z
1: TOF MS ES+
1.21e6
17540
Calc. desired mass: 17539 Da
Obs. mass: 17540 Da
mass

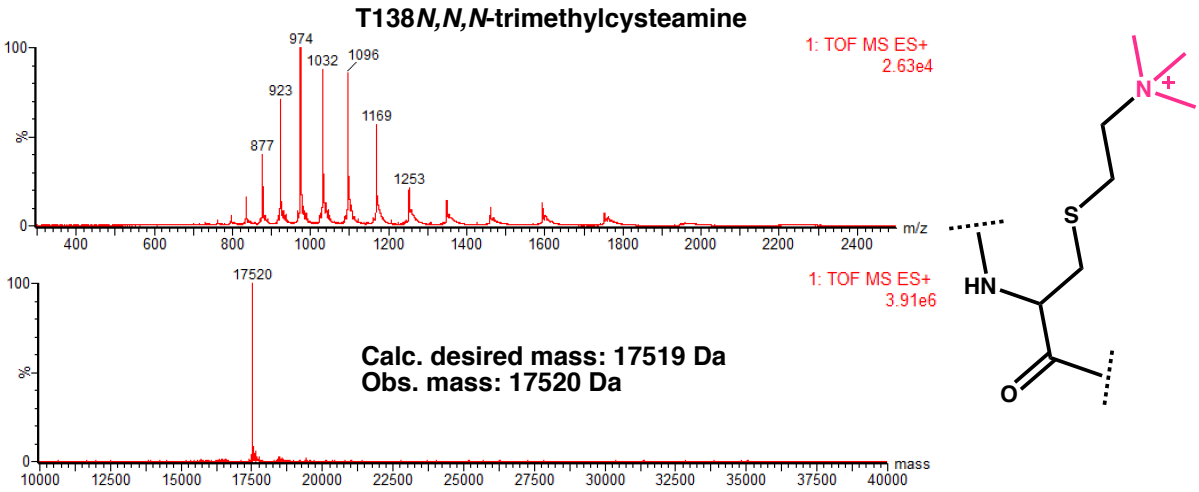
c


T138*N*-acetylcysteamine
1: TOF MS ES+
2.56e4
974
923
1031
1096
877
835
797
763
1169
1252
1348
m/z
1: TOF MS ES+
4.88e6
17517
Calc. desired mass: 17517 Da
Obs. mass: 17517 Da
mass

d



e



f

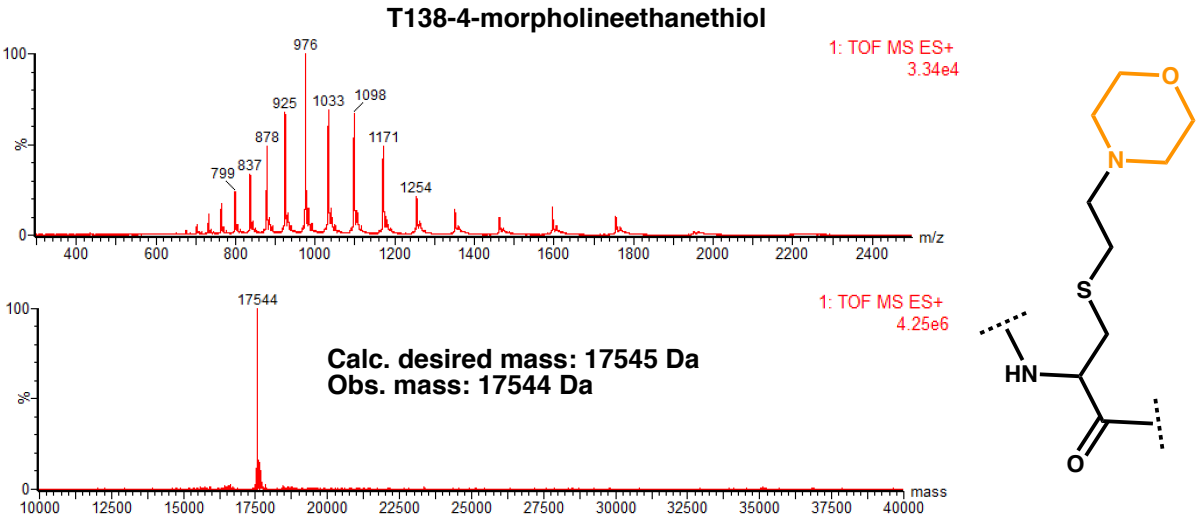


Figure 2.21. Cysteamine derivative reactions. (a) Panel of cysteamine derivatives for activity screening. (b) LC-MS data for T138MESNA confirming a homogeneous conversion. (c) LC-

MS data for T138Acetyl-cysteamine confirming a homogeneous conversion. (d) LC-MS data for T138Captamine confirming a homogeneous conversion. (e) LC-MS data for T138*N,N,N*-trimethylcysteamine confirming a homogeneous conversion. (f) LC-MS data for T138-4-morpholineethanethiol confirming a homogeneous conversion.

As with all of the previous chemical mutants, these cysteamine derivates had their structural integrity assessed by CD. This revealed that once again all of the chemical mutants displayed a similar profile to the starting scaffold (**Figure 2.22**).

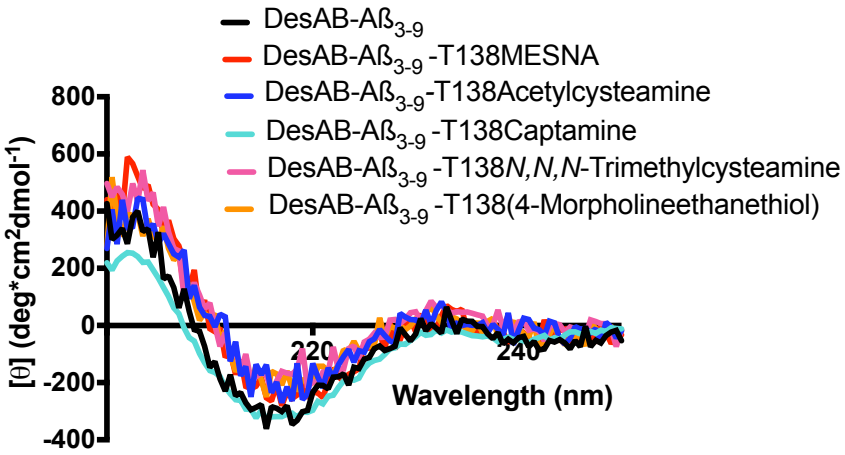
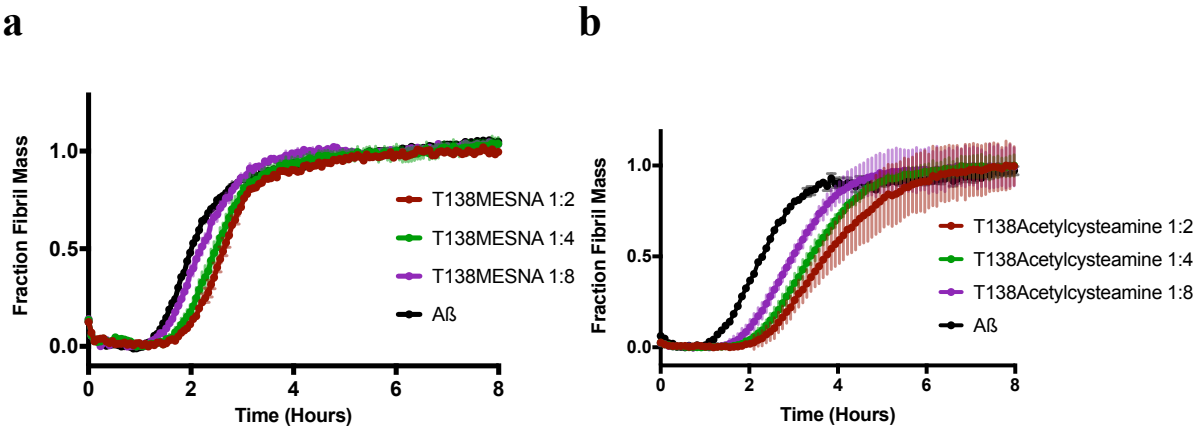


Figure 2.22. Cysteamine derivative chemical mutant's CD profiles. CD traces of DesAB-A$\beta_{3\text{-}9}$ and the various cysteamine derivatised chemical mutants at position T138.

The derivatives were then screened for their anti-aggregation activity in the same manner as previously described. This second screen at position 138 found that T138Captamine had an even more potent inhibition of Aβ aggregation compared to the starting activity and to T138Cysteamine (**Figure 2.23a-e**).



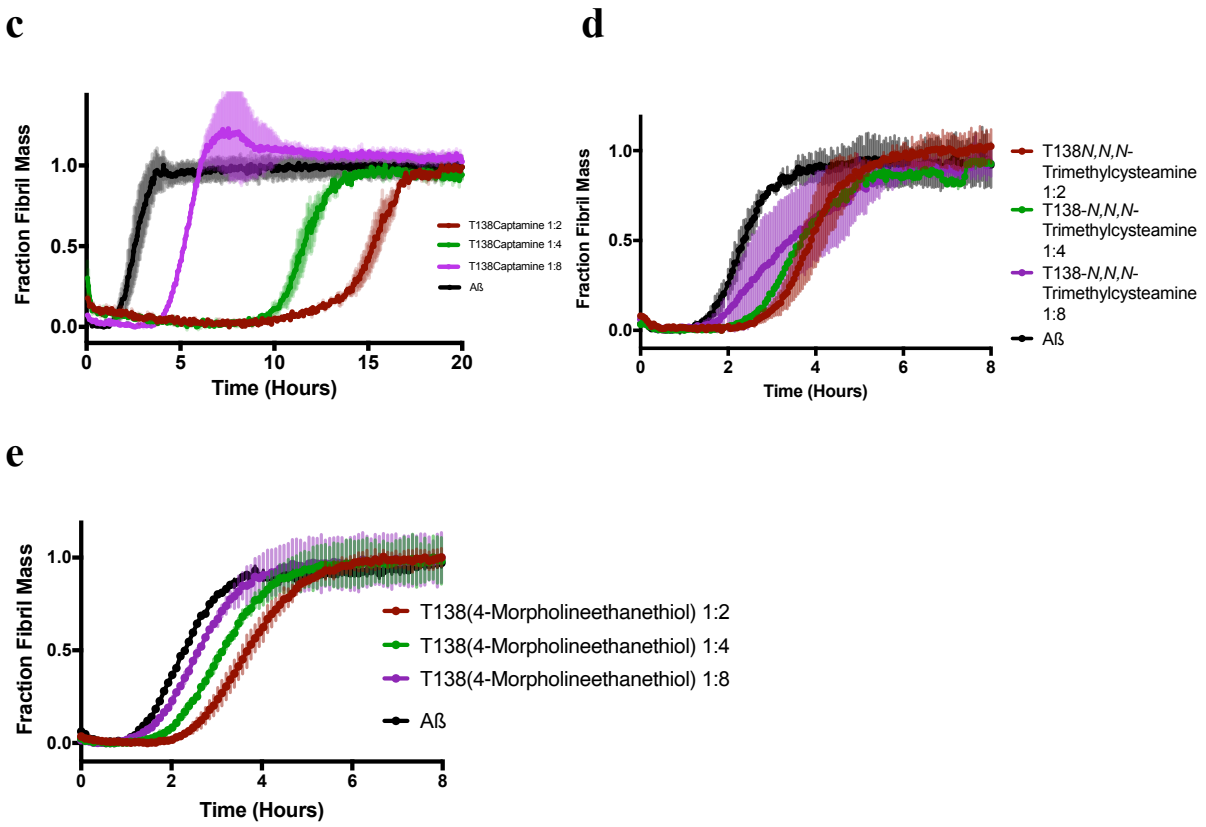


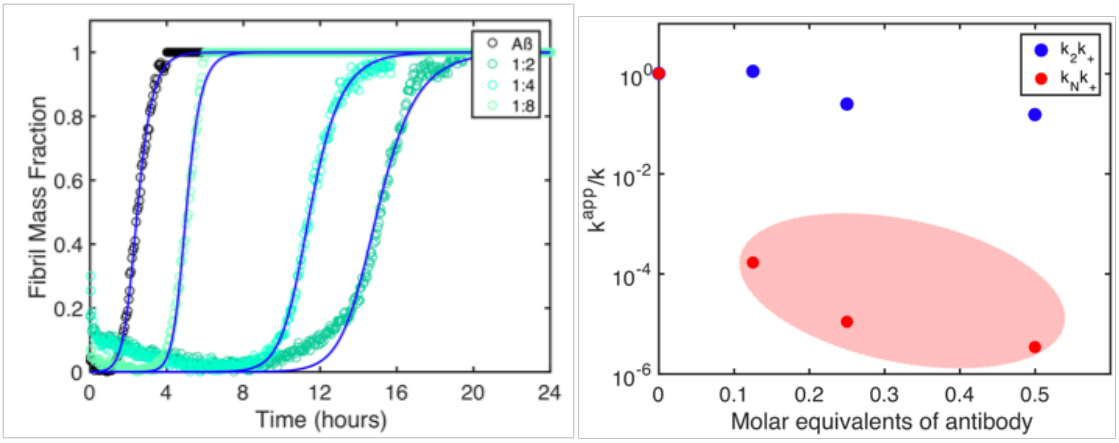
Figure 2.23. T138Captamine has an enhanced inhibition of aggregation. (a-e) Aggregation curves for cysteamine derivatives at position T138. DesAB-Aβ$_{3-9}$ (T138Captamine) in panel (c) is the only one that displays a significant improvement compared to T138Cysteamine. The Aβ concentration for each reaction is 1.5 μM and the antibody concentration in the 1:2 condition is 750 nM, in the 1:4 condition the concentration is 375 nM, and in the 1:8 condition it is 187.5 nM.

To confirm which microscopic step in the aggregation reaction the captamine derivative enhanced a kinetic analysis was once again performed. This analysis confirmed that the captamine mutation built upon the previous potentiation of the primary nucleation reaction that cysteamine had over the starting construct (**Figure 2.24a,b**). When the primary nucleation rates of the 1:2 dilutions of the starting DesAB-Aβ$_{3-9}$ and the T138Captamine chemical mutant are directly compared it can be seen that there is a nearly 5-order of magnitude decrease in the nucleation rate global parameter with this single amino-acid change (**Figure 2.24c**). This 5-order of magnitude decrease in the rate constant, however, only leads to a ~10-fold retardation of the half-life at the 1:2 condition. This is due to the unperturbed contributions of the secondary nucleation pathways. To compare this potency to the activity of different anti-aggregation agents, one parameter to look at is the inhibitory concentration that yields a 50% increase in the time to reach peak fibril mass (known as the KIC$_{50}$). A recent development campaign to yield small molecule inhibitors of Aβ aggregation produced candidates

approaching a KIC_{50} of 1 μM[89]. Our lowest tested concentration of 187.5 nM for DesAB-$Aβ_{3-9}$ (T138Captamine) yielded a ~200% increase in time to peak fibril mass. Therefore, it is likely that the KIC_{50} is in the range of 100-150 nM; nearly 10 times the potency of the best small molecules. However, more detailed experiments in the lower concentration ranges will need to be conducted to definitively determine the KIC_{50}.

At this point, the primary nucleation parameter that has been used for comparison has not been decoupled from the elongation rate ($k_N k_+$, where k_+ is the elongation rate). To be sure that the majority of the inhibition can be directly attributed to the inhibition of primary nucleation, rather than elongation, we performed an inhibition screen where elongation is the dominant form of fibril growth. This was achieved by performing the aggregation reaction in the presence of 30% (by monomer concentration) preformed Aβ fibrils. With this high concentration of fibrils, the contribution of primary and secondary nucleation to fibril growth can be considered negligible[85]. So, in this case with 1.5 μM Aβ monomer, 450 nM fibrils were added, and the antibodies were added in a 1:1 ratio to monomer. This revealed that the T138Captamine chemical mutant had only a slight effect on inhibiting fibril elongation compared to the starting scaffold, confirming that most of the inhibitory activity can be directly attributed to the inhibition of primary nucleation (**Figure 2.24d**).

a **b**



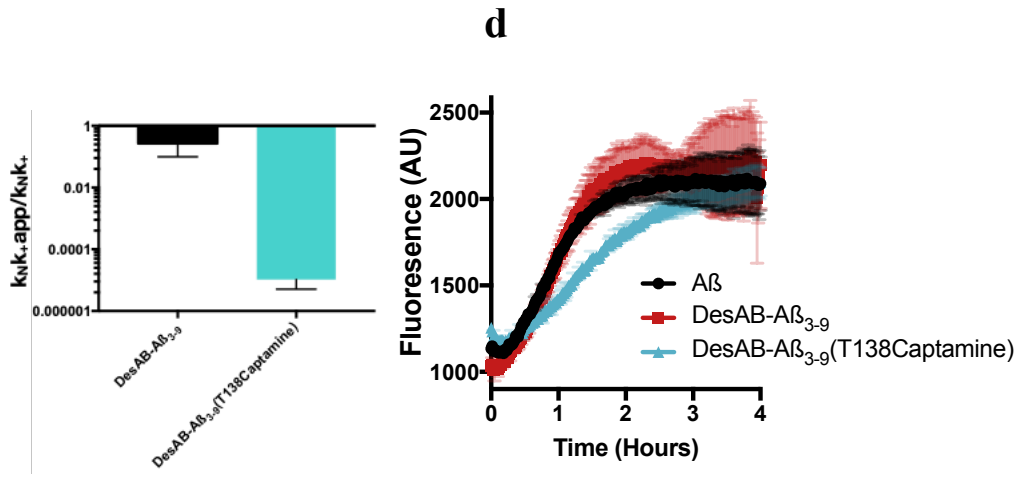


Figure 2.24 Kinetic characterisation of DesAB-Aβ3-9(T138Captamine). (a) Fits of the kinetic model to the dilution series of the anti-aggregation screen. (b) Decoupled global parameters for the primary and secondary nucleation rates, showing that the captamine mutation further potentiates the inhibition of primary nucleation. (c) Comparison of the apparent primary nucleation rates for the 1:2 antibody dilutions of DesAB-Aβ3-9 (black) and T138Captamine (teal). (d) Screen for the inhibition of fibril elongation by DesAB-Aβ3-9 (red) and T138Captamine (teal).

We then characterized the binding of DesAB-Aβ(3-9)(T138Captamine) and the starting construct to C-terminally immobilized monomeric Aβ42 by biolayer interferometry (BLI). The C-terminally biotinylated Aβ was loaded onto streptavidin BLI tips overnight at 5 ºC while control streptavidin tips were coated with biocytin. The association and dissociation signals of each the antibodies to the control tips was subtracted from their respective signals produced by the Aβ coated tips to account for any non-specific binding. The results revealed that the two antibodies bound monomeric Aβ42 under these conditions in a very similar manner (**Figure 2.25a**). We note that, since it is difficult to apply BLI to accurately quantify the binding to Aβ42, in particular because immobilizing Aβ42 drastically alters the structural ensemble that this disordered peptide adopts compared to its free state in solution, these data do not conclusively explain the mechanism behind the enhanced inhibition, which could also be brought about by the enhanced association with primary oligomers or other aggregated species.

Finally, we measured the conformational stability of our potentiated single-domain antibody DesAb-Aβ(3-9)(T138captamine) by means of thermal denaturation. Our results show that the melting temperature is identical to that of DesAb-Aβ(3-9) within the measurement error, and the melting profiles essentially overlap (**Figure 2.25b**). This observation is particularly

interesting because through conventional methods of activity maturation, there is often a significant trade-off between stability and activity when comparing the starting and end constructs, a phenomenon that has been reported to be particularly extreme in the case of aggregation-prone antigens like Aβ[116]. The single uniform melting curve also supports the integrity of the disulfide bond in both the chemical mutant and starting single-domain antibody, as a reduced population yields a multistep melting curve.

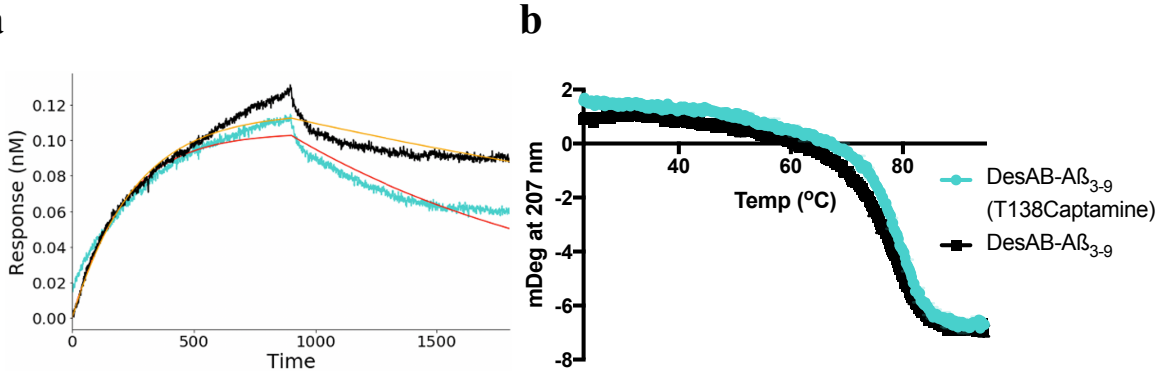


Figure 2.25 BLI and thermal stability comparisons. (a) BLI association and dissociation curves for DesAB- $A\beta_{3\text{-}9}$ (black, orange fit) and T138Captamine (teal, red fit) to C-terminally immobilised $A\beta_{42}$. (b) Thermal denaturation curves measured by CD at 207 nM for DesAB-$A\beta_{3\text{-}9}$ (black) and T138Captamine (teal).

2.3 Conclusion

In this chapter we investigated if chemical mutagenesis could be used to conduct a structure activity relationship beyond the 20 canonical amino acids on a single-domain antibody. Our starting antibody had an initially nascent designed activity to inhibit the aggregation of the Aβ peptide. The first step to approach this project was to install the synthetically versatile non-canonical amino acid Dha at each position along the CDR3 loop. This was accomplished by initially installing cysteine at each position in the loop by genetic means then eliminating the free thiol through a bis-alkylation-elimination reaction to Dha post-expression. Encouragingly for this approach, the conversion to Dha was uniformly successful along the loop, potentially meaning most CDR sequences could be amenable to this chemical activity maturation approach. The next course of action was to screen Dha at each position for its own accessibility for modification. To investigate this, a reactivity screen was performed with the benchmark reagent β-mercaptoethanol where the time to completion was closely monitored by LC-MS. This experiment revealed that the local chemical environment surrounding Dha has a powerful influence over its reactivity, with some positions reaching completion in <15 minutes while others took as long as 6 hrs. Moving forward with the three

most accessible sites for side-chain exploration, we attempted two other routes to modification other than thia-Michael additions: free radical based C-C bond formation and aza-Michael ligation. However, neither of these two approaches were pursued after an initial investigation due to poor reaction efficiency and detrimental effects on the protein structure, respectively. Using solely thiol-based reagents, initial side-chain exploration was conducted at the 3 sites with diverse functional groups. Subsequent activity screening for the enhanced inhibition of Aβ aggregation revealed that one chemical mutant (T138Cysteamine) had improved inhibition of the primary nucleation reaction. In the spirit of small molecule SAR studies, the initial cysteamine "hit" was diversified with subtle to dramatic chemical changes to try to drive the potency even more. This identified that the addition of two methyl groups to the terminal amine further drove the potency for inhibiting the primary nucleation reaction, ending up 5 orders of magnitude more potent than the starting sequence. Importantly, this approach of chemical activity maturation led to no compromise in the stability of the protein. Activity and stability trade-offs are commonly encountered in traditional directed evolution campaigns[112], so this type of hyper-specific chemical modification to single amino acids to drive activity may be a desirable course of action when other biophysical traits must be safeguarded.

Chapter 3: Facile installation of tau PTM mimetics via chemical mutagenesis

3.1 Introduction

3.1.1. Tau PTMs and their study

Many crucial roles for tau in neuronal function are becoming recognized beyond regulating microtubule dynamics[64]. This expanding diversity of function is no doubt a consequence of the incredible diversity and shear number of PTMs that tau has been observed to undergo[83]. A plethora of approaches have been taken to accurately study the effects of tau PTMs. Early on in the literature, the effects of phosphorylation at specific sites was investigated by genetic means through the mutation of Ser/Thr residues to Glu/Asp to mimic the physiochemical changes brought about by phosphorylation[170]. While this approach proved initially useful, the terminal carboxylic acid of glutamic or aspartic acid is ultimately a poor approximation of phosphorylation for a number of reasons. The pK_a value for the carboxylate side chains are ~ 4 while the first pK_a value for phosphate is ~ 2 and the second ~ 6. This means that at physiological pH a Glu/Asp residue will only have a -1 charge while phosphate will be -2. The geometry is also inconsistent, with the carboxylate group in a trigonal planar structure while phosphate groups are tetrahedral. Similar physiochemical discrepancies can also be listed for genetic mimics of other PTMs like using Gln/Asn to mimic acetylation. Purified tau can be phosphorylated *in vitro* when combined with specific kinases under appropriate conditions, however, this approach usually leads to a heterogeneous mixture of sites being modified[171,172]. Therefore, making it difficult to accurately assign effects to single specific modifications. Enzymatic modification *in vitro* is also only possible for certain observed modification of tau. These inescapable short comings of genetic approaches have led to the development of more sophisticated approaches to accurately study protein PTMs as a whole.

A particular approach of note for tau is a semi-synthesis protocol developed by Lashuel and co-workers used to site-specifically incorporate exact PTMs into tau[173] (**Figure 3.1**). This is a particularly powerful approach as it allows more than 1 PTM to be incorporated at a time and can introduce disparate and non-enzymatic PTMs altogether. The protocol calls for a large N-terminal fragment of tau to be recombinantly expressed and purified from bacteria, then

shorter 50-60 amino acid fragments of the C-terminus are created through solid phase peptide synthesis and combined via native chemical ligation, taking advantage of the two endogenous Cys residues. While the power of this approach is clear, this protocol can only be carried out by an experienced synthetic chemist specialising in peptide chemistry, a type of person not readily available to most labs interested in studying PTM effects on tau. It is therefore desirable to develop more universally friendly approaches to generate site-specific PTMs on tau.

Here chemical mutagenesis can again prove its utility. Some of the earliest applications of chemical mutagenesis in its resurgence were to accurately study PTMs. Different approaches of chemical mutagenesis have been used to study kinase activation, histone modifications, and incredibly complex glycosylation patterns to list a few examples[163,160,174]. From the successes of these diverse projects, tau appears ripe for the application of this technique.

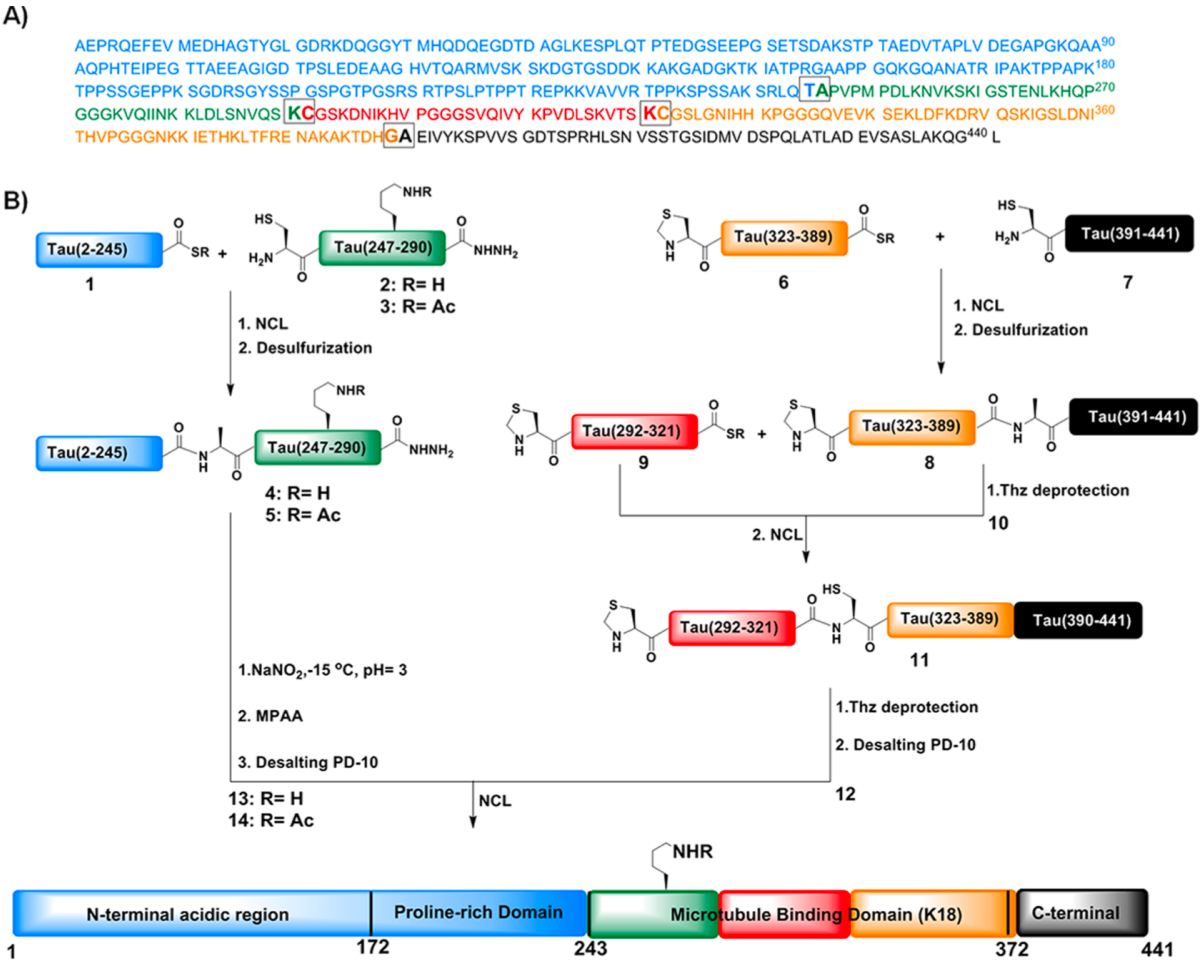


Figure 3.1. Semi-synthesis of Tau bearing PTMs. (a) Sequence of tau broken down into regions for semi-synthesis. (b) Semi-synthesis scheme. Figure adapted with permission from reference [173].

3.1.2. Chapter aims

The overall aim of this chapter will be to assess the amenability of tau to create several PTMs at different sites along this large protein. Since phosphorylation is the most common and well-studied modification of tau, three disparate sites of known phosphorylation will be explored as well as a dual phosphorylation. Two of these sites, S262 and S356, have been previously established to significantly impair tubulin polymerisation activity[171,175]. These two sites will serve to verify the accuracy of the PTM mimetics installed via Dha. These sites will also be the two pursued in the dual phosphorylation reaction. The other site, S199 in the proline rich domain (PRD), has been shown to be a biomarker for AD[174,175], but has not had the actual effects of phosphorylation on the protein behaviour explored. If the S262 and S356 positions verify the accuracy of this approach, to our knowledge this work will constitute the first characterisation of phospho-S199 on tau function. The growing importance of modifications other than phosphorylation on Tau cannot be ignored. To this end, we will also look to install lysine acetylation and dimethylation mimetics at K311, located in one of the most amyloidogenic portions of the protein[68]. The acetylation mimic will also serve to verify the accuracy of this approach, as Lashuel and co-workers have characterised acetylation at K280 via their semi-synthesis approach, which is located in a nearly identical repeat earlier on in the protein sequence[173]. Dimethylation on K311, however, has not had its effects explored yet and has only been relatively recently characterised in proteomics experiments from tissue samples[83]. The first aim for all of these positions and modifications will be to purify the respective cysteine mutants, second to find an approach to efficiently convert these to Dha, then to develop the reactions for the final modifications, and lastly to perform an initial characterisation of these PTMs by assessing their effects on the tubulin polymerisation activity of tau. Some of the work in this chapter was done with the assistance of Ross Taylor during his BBSRC rotation in the Vendruscolo lab from October 2019 to December 2019.

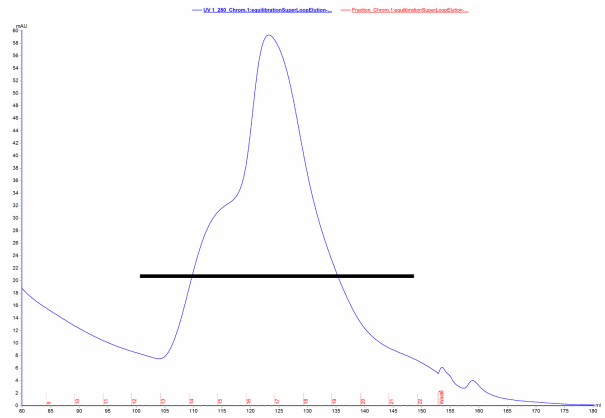
3.2 Results

3.2.1. Tau mutagenesis, expression and purification

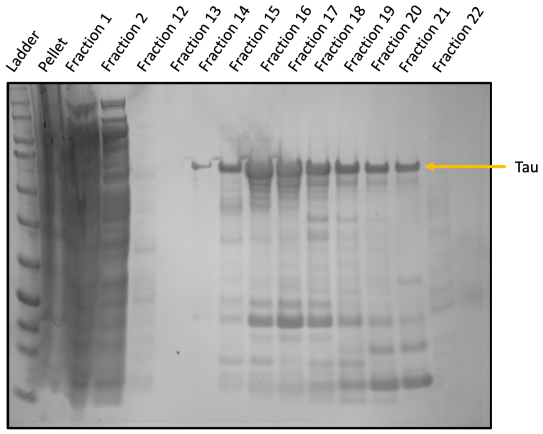
Recombinant 2N4R tau was expressed in a similar manner to previously established[178]. The original pet29b vector containing the 2N4R (C291S & C322S) sequence was gratefully gifted by the Klenerman lab. Site-directed mutagenesis was carried out on this plasmid in the

same manner as the antibodies in chapter 3 to create the desired cysteine mutants (S199C, S262C, S356C, S262C & S356C, and K311C) (see Materials and Methods)[169]. Once the sequences were verified, the plasmids were transformed into BL21 Gold DE3 cells for expression. The various mutants were expressed in 6x1 L cultures overnight at 18 ºC after induction with IPTG. The bacteria were then harvested by centrifugation and the pelleted cells were resuspended in 50 mM MES (pH 6.5), 5 mM DTT, 0.1 mM PMSF and lysed via sonication. The lysed solution was then centrifuged again, and the supernatant was collected and kept. After the addition of RNAse and DNAse, the supernatant was passed through a filter before undergoing cation exchange chromatography using a Hitrap CaptoS column. All the fractions in the highest UV signal at the end of the salt gradient were collected and assessed by SDS-PAGE for the presence of Tau (**Figure 3.2a,b**). All fractions containing Tau were pooled and precipitated on ice overnight by the addition of 20% (w/v) ammonium sulphate. The precipitated solutions were pelleted via centrifugation and then resuspended in a small amount of SSPE buffer containing 5 mM DTT, this solution was then run on a Superdex 200 Increase 10/300 GL SEC column to isolate the pure tau, fractions under the appropriate UV signal were again assessed by SDS-PAGE and only the purest fraction was kept for experiments (**Figure 3.2c-i**). All variants were expressed and purified in good yield.

a



b



c **d**

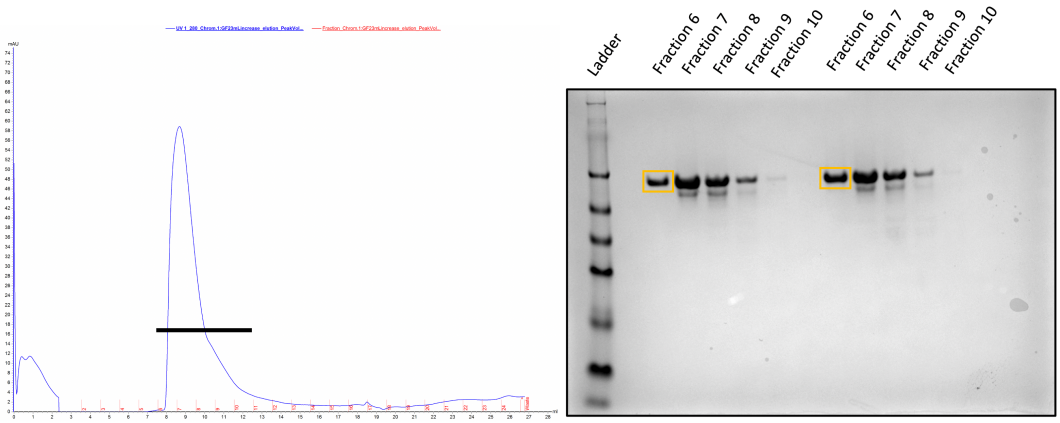


Figure 3.2. Tau purification. (a) Representative chromatogram from the cation exchange step of 2N4R tau, black bar indicated the collected fractions. (b) Image of a representative SDS-PAGE of the collected fractions from the cation exchange (along with the pellet and early on fractions), yellow arrow points to tau. (c) Chromatogram from a representative SEC of precipitated tau, black bar again indicates collected fractions. (d) SDS-PAGE of the fractions collected from two different SEC runs (WT Tau and S356C Tau, respectively), fractions with the yellow box indicate those kept for experiments.

3.2.2. Dha formation on Tau

The same approach to Dha formation that was taken in **Chapter 2** was also applied to tau. The first cysteine variant that was purified was S199C, and so the reaction with DBHDA was first attempted with this protein. S199C was buffer exchanged into 20 mM NaP_i (pH 8) and then 500x molar equivalents of DBHDA was added and then allowed to react at 37 ºC and 500 rpm, exactly like the antibodies. Unfortunately, checking the reaction by LC-MS revealed that even after very extensive incubations of 24 hrs the protein was stuck at a stable intermediate along the pathway to forming Dha (**Figure 3.3**). These extended incubations also led to very noisy spectra. From these preliminary reactions, it was clear that a more efficient approach to forming Dha on tau is needed. This also serves to highlight that any chemical mutagenesis approach must be heavily tailored to the protein of interest, which is not commonly communicated in the literature of the field.

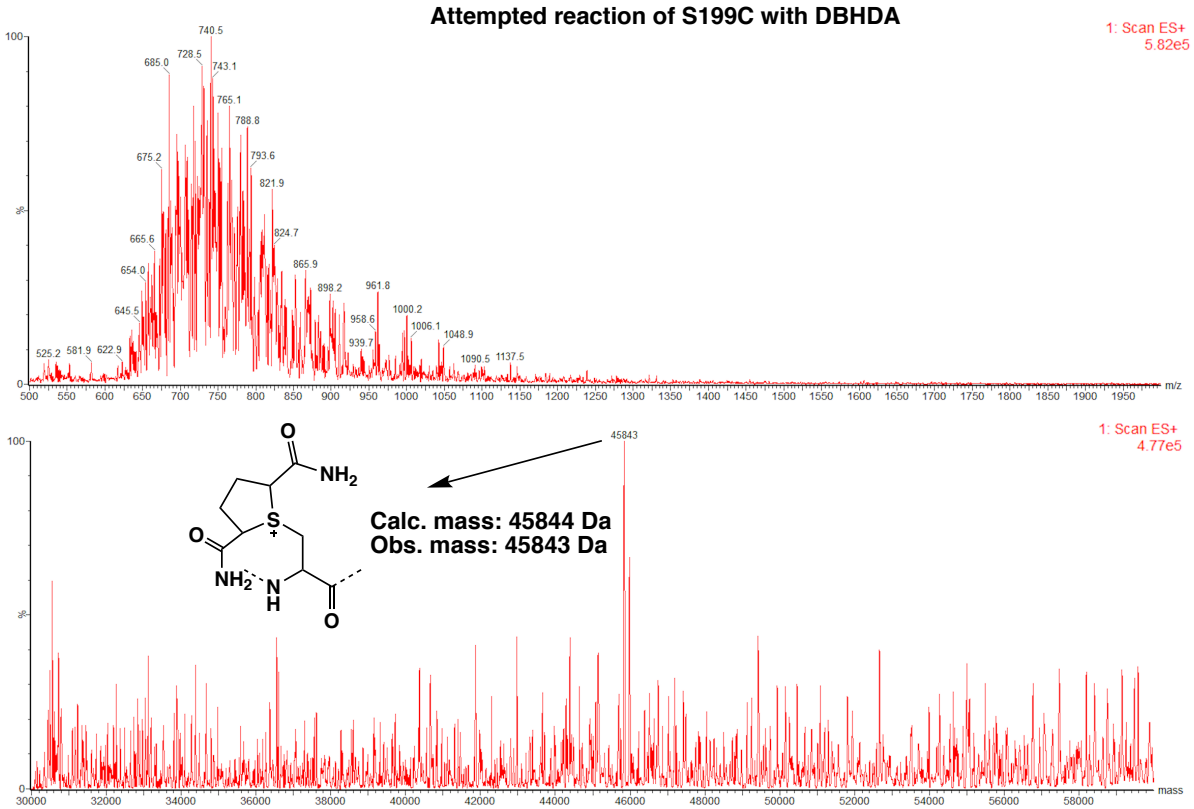
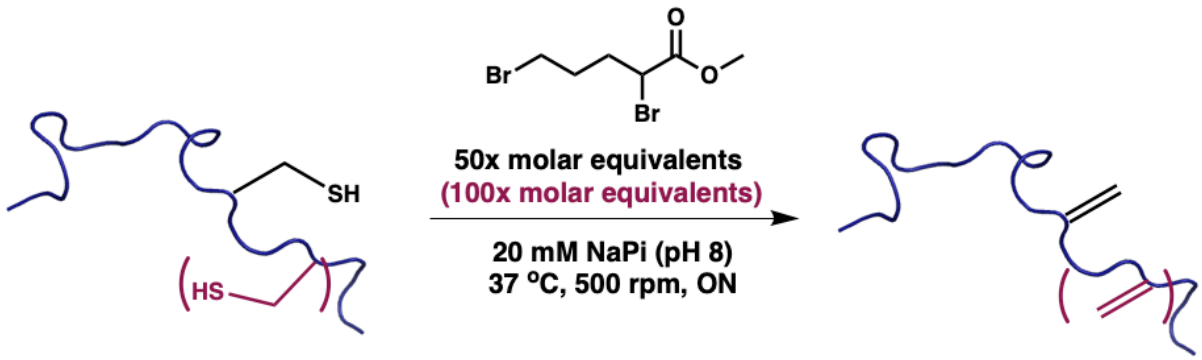


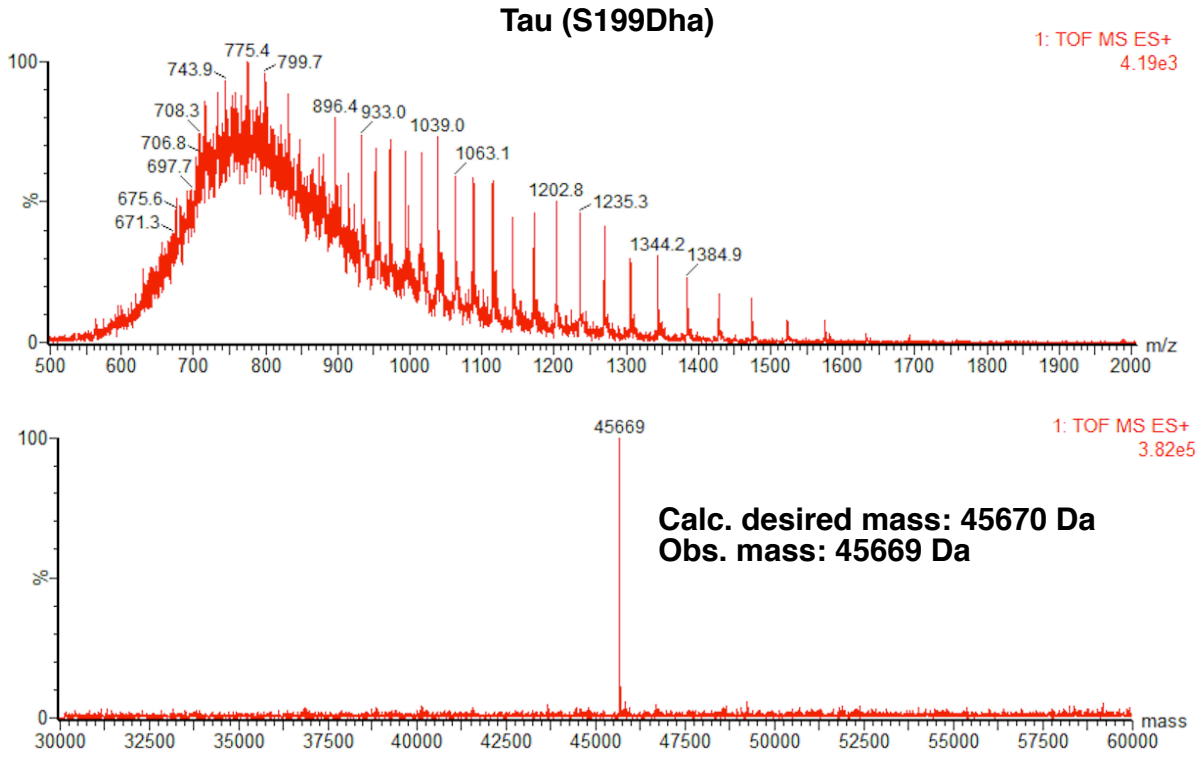
Figure 3.3. Attempted Dha formation on Tau via DBHDA at 24hrs. LC-MS spectrum of the reaction of S199C with DBHDA after 24 hrs using the same conditions as the antibodies. It can be seen that the reaction has stalled at one of the intermediates on the way to Dha.

Luckily for us, a more efficient bis-alkylation elimination reagent has recently been developed[179]. The reagent, methyl 2,5-dibromopentanoate (MDBP) (**Figure 3.4a**), is a slightly water miscible liquid that has a much-enhanced rate of both addition to cysteine and subsequent elimination compared to DBHDA[162]. This molecule was engineered to allow for the conversion of multiple Cys residues to Dha on peptides, avoiding the unwanted side product of an intramolecular addition of an unconverted Cys to Dha elsewhere on the protein, which will be very important for the dual modification of the S262C & S356C mutant. Using the same buffer and conditions as previously for DBHDA, 50 molar equivalents of MDBP was added to each of the mutants and allowed to react at 37 ºC shaking at 500 rpm overnight. Checking the reactions by LC-MS in the morning confirmed that the MDBP regent afforded full conversions to Dha, even for the dual S262C & S356C construct (**Figure 3.4b-f**).

a

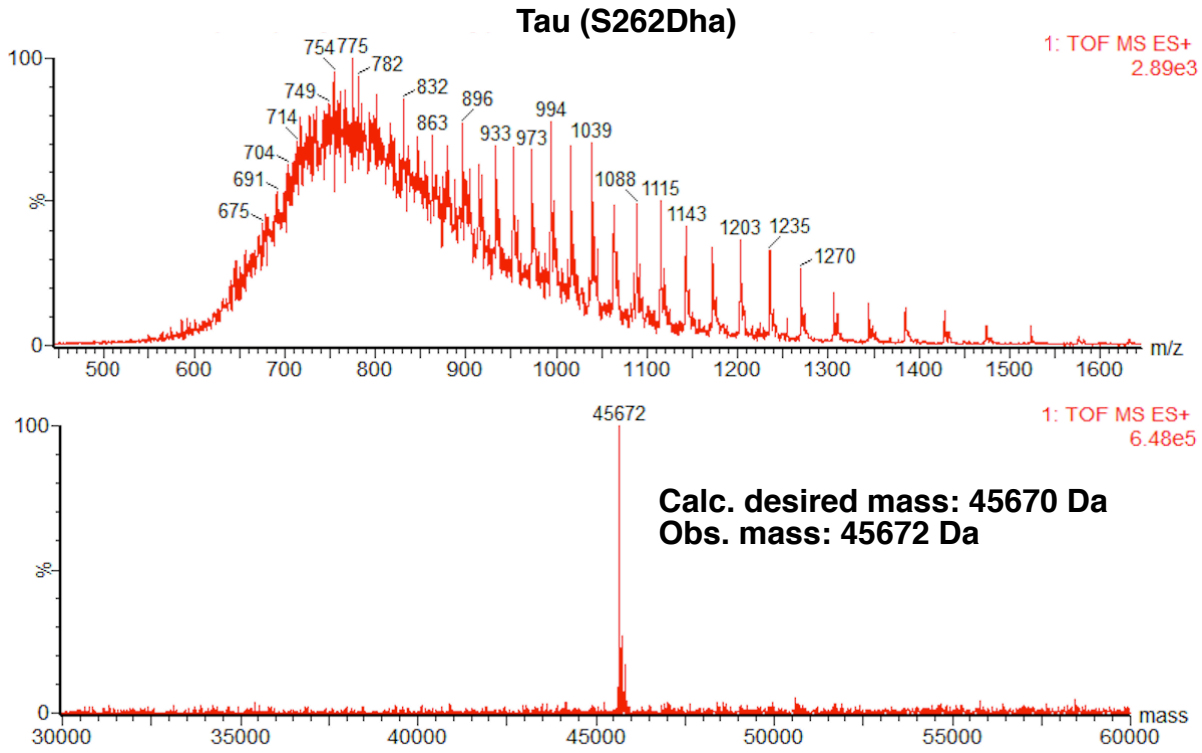

Br
O
O
Br
SH
HS
50x molar equivalents
(100x molar equivalents)
20 mM NaPi (pH 8)
37 °C, 500 rpm, ON

b

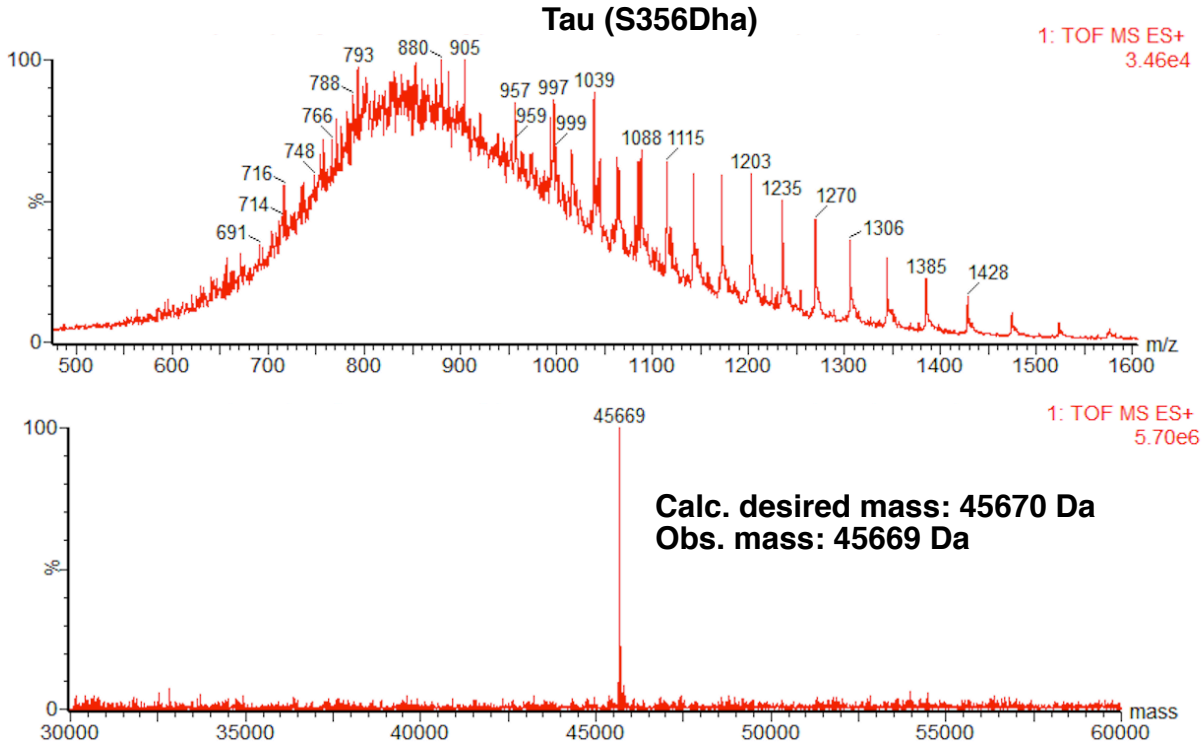

Tau (S199Dha)
1: TOF MS ES+
4.19e3
100
%
0
775.4
743.9
799.7
708.3
706.8
697.7
675.6
671.3
896.4
933.0
1039.0
1063.1
1202.8
1235.3
1344.2
1384.9
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
m/z

1: TOF MS ES+
3.82e5
100
%
0
45669
Calc. desired mass: 45670 Da
Obs. mass: 45669 Da
30000
32500
35000
37500
40000
42500
45000
47500
50000
52500
55000
57500
60000
mass

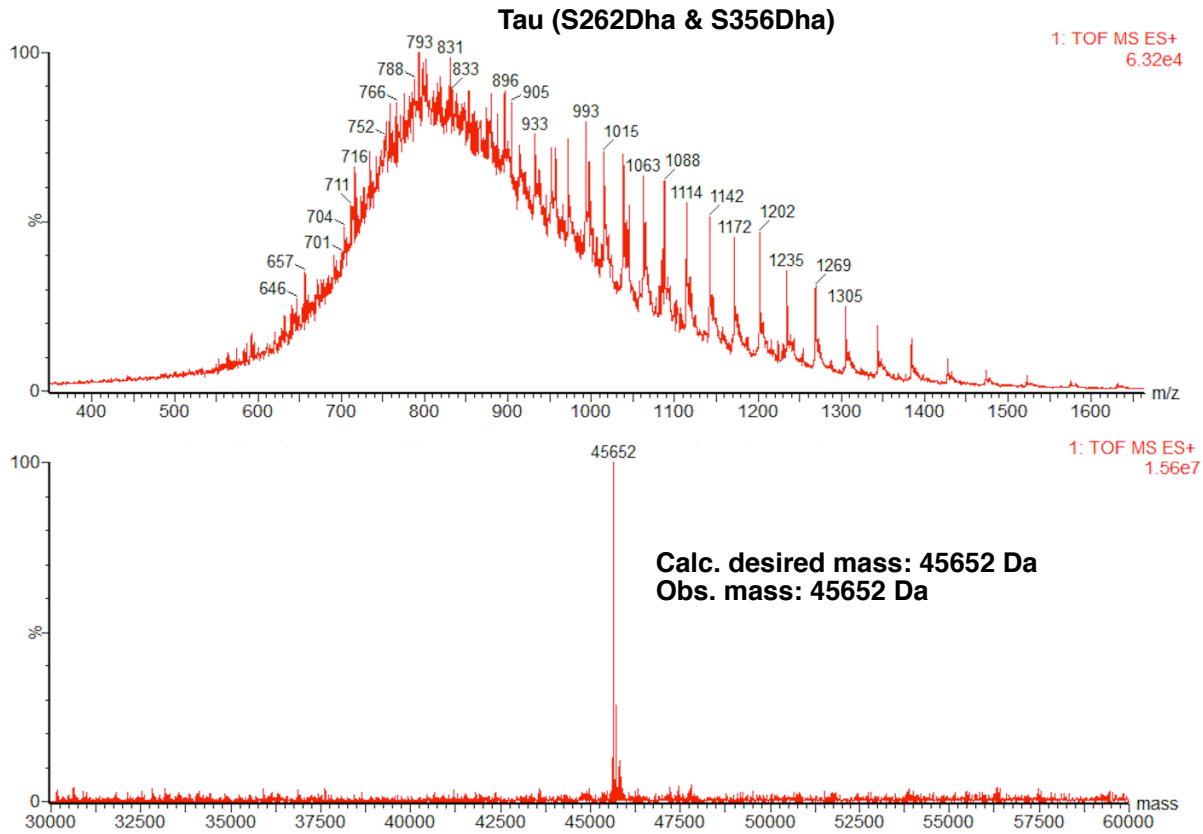
c


Tau (S262Dha)
1: TOF MS ES+
2.89e3
100
%
0
754 775
782
749
832
714
896
863
994
933 973
1039
704
691
1088
1115
675
1143
1203 1235
1270
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
m/z
45672
1: TOF MS ES+
6.48e5
Calc. desired mass: 45670 Da
Obs. mass: 45672 Da
30000 35000 40000 45000 50000 55000 60000
mass

d


Tau (S356Dha)
1: TOF MS ES+
3.46e4
100
%
0
793
880
905
788
957 997
1039
766
959
999
748
1088 1115
716
1203
714
1235
1270
691
1306
1385
1428
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
m/z
45669
1: TOF MS ES+
5.70e6
Calc. desired mass: 45670 Da
Obs. mass: 45669 Da
30000 35000 40000 45000 50000 55000 60000
mass

e


Tau (S262Dha & S356Dha)
1: TOF MS ES+
6.32e4
100
%
0
793
831
788
833
766
896
905
752
933
993
716
1015
711
1063
1088
704
1114
1142
701
1172
1202
657
1235
1269
646
1305
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
m/z
1: TOF MS ES+
1.56e7
45652
Calc. desired mass: 45652 Da
Obs. mass: 45652 Da
30000
32500
35000
37500
40000
42500
45000
47500
50000
52500
55000
57500
60000
mass

f

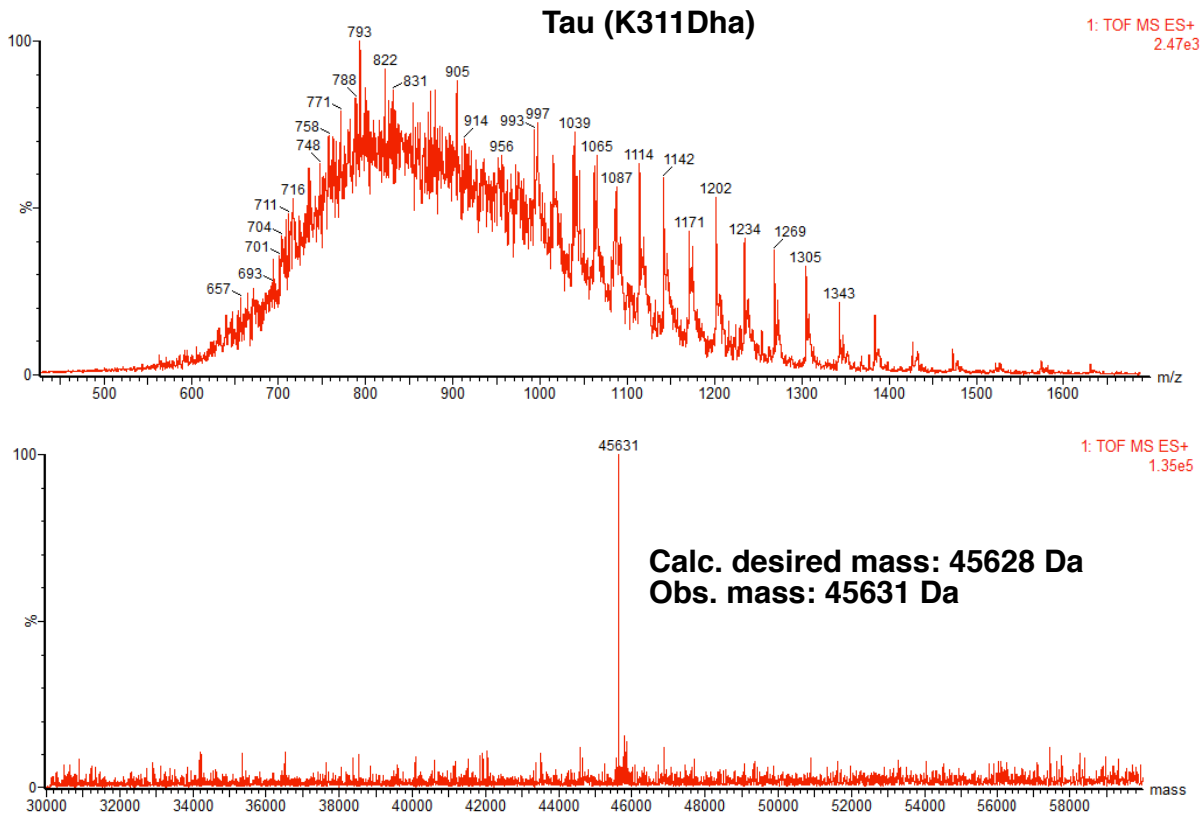
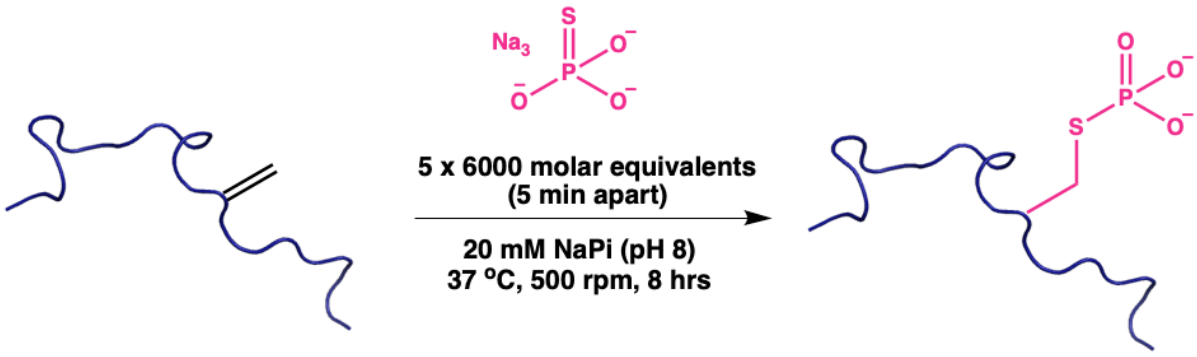

Tau (K311Dha)
1: TOF MS ES+
2.47e3
100
%
0
793
822
788
831
905
771
914
993
997
758
1039
748
956
1065
1114
1142
716
1087
711
1202
704
1171
1234
1269
701
1305
693
1343
657
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
m/z
1: TOF MS ES+
1.35e5
45631
Calc. desired mass: 45628 Da
Obs. mass: 45631 Da
30000
32000
34000
36000
38000
40000
42000
44000
46000
48000
50000
52000
54000
56000
58000
mass

Figure 3.4. Dha formation on Tau with MDBP. (a) Reaction scheme for Cys containing tau reactions with MDBP, maroon writing and side chains are for the dual S262C & S356C mutant. (b) LC-MS post reaction confirming Dha formation at position 199. (c) LC-MS post reaction confirming Dha formation at position 262. (d) LC-MS post reaction confirming Dha formation at position 356. (e) LC-MS post reaction confirming Dha formation at both 262 and 356 positions in the duel mutant. (f) LC-MS post reaction confirming Dha formation at site 311.

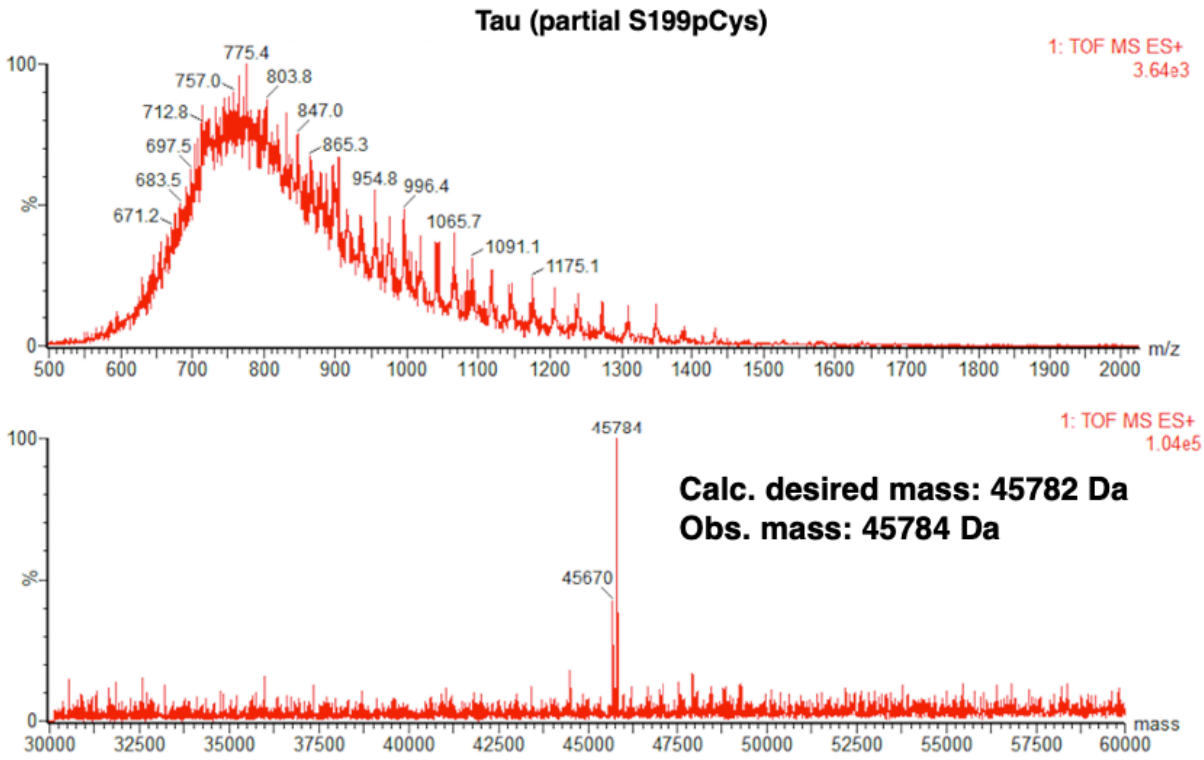
3.2.3. Single phosphorylation reactions at Dha

In order to install an accurate serine phosphorylation mimic at Dha, the protein must be reacted with sodium thiophosphate (**Figure 3.5a**). This reaction installs the phospho-amino acid phosphocysteine (pCys), which recapitulates all of the physiochemical properties of phosphorylation except for the P-S-C connection versus P-O-C. Before the reaction begins, the sodium thiophosphate needs to be prepared by mixing in an appropriate amount of HCl in order to bring the pH to approximately 8 (see **Materials and Methods**). Once this was complete, the sodium thiophosphate solution was slowly added to 100 µLs of Dha bearing tau (S199Dha, S262Dha, and S356Dha) in reaction buffer at 50 µM and allowed to react at 37 °C and 500 rpm. After 4 hours the reaction was checked by LC-MS and revealed that it had only progressed to approximately 75% completion uniformly at the three positions (**Figure 3.5b**). Extending the duration of the reaction to 8 hrs afforded >95% conversion at all three sites (**Figure 3.5c-e**).

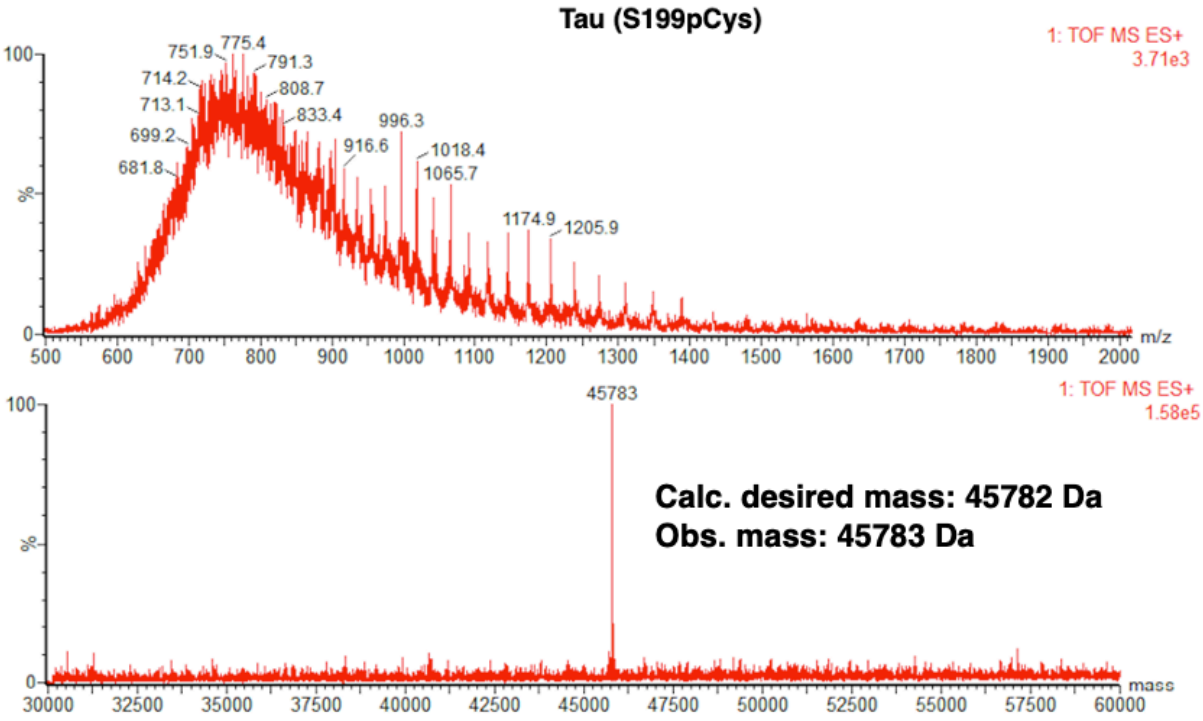
a



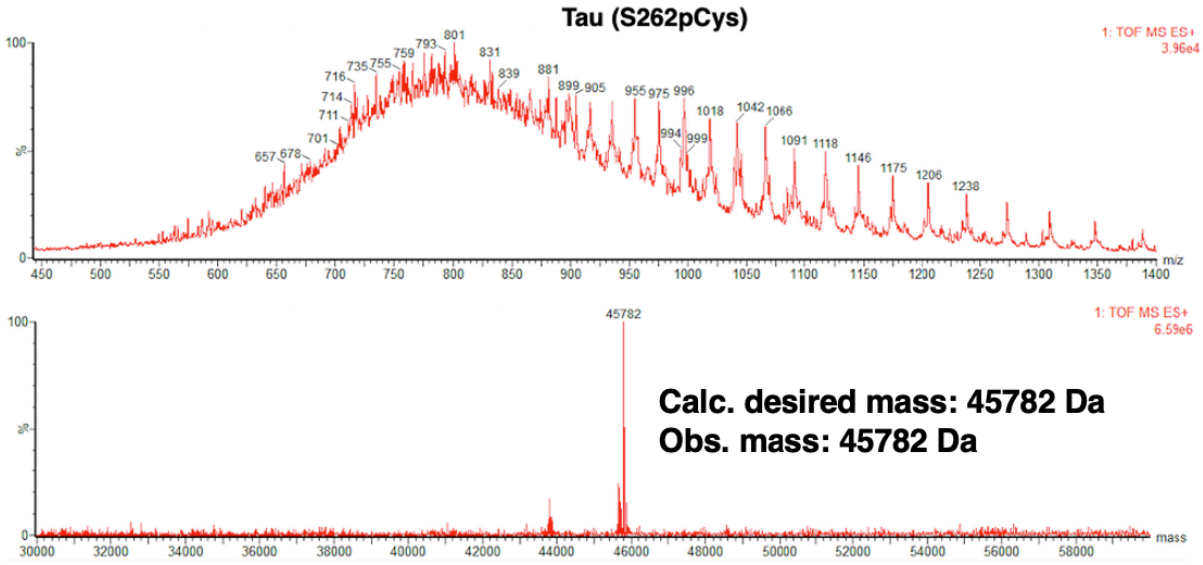
b


Tau (partial S199pCys)
1: TOF MS ES+
3.64e3
775.4
757.0
803.8
712.8
847.0
697.5
865.3
683.5
954.8
996.4
671.2
1065.7
1091.1
1175.1
m/z
1: TOF MS ES+
1.04e5
45784
Calc. desired mass: 45782 Da
Obs. mass: 45784 Da
45670
mass

c


Tau (S199pCys)
1: TOF MS ES+
3.71e3
751.9
775.4
791.3
714.2
808.7
713.1
833.4
996.3
699.2
916.6
1018.4
681.8
1065.7
1174.9
1205.9
m/z
1: TOF MS ES+
1.58e5
45783
Calc. desired mass: 45782 Da
Obs. mass: 45783 Da
mass

d



e

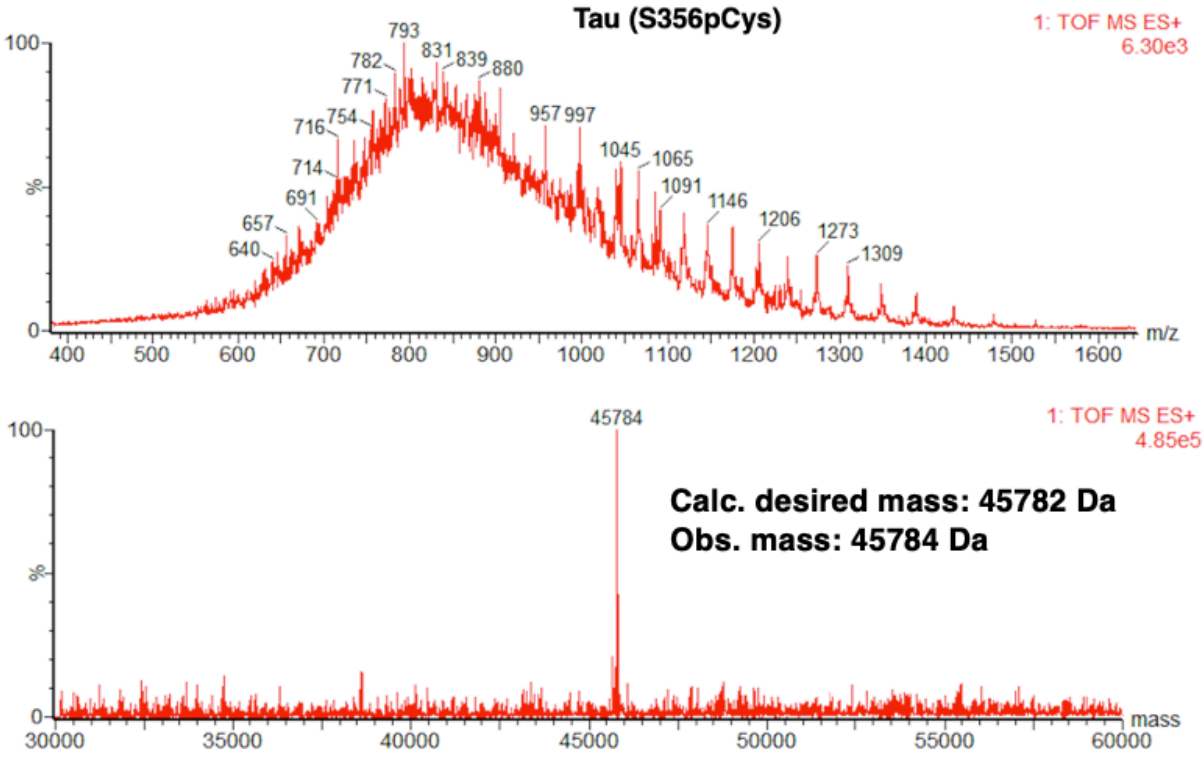


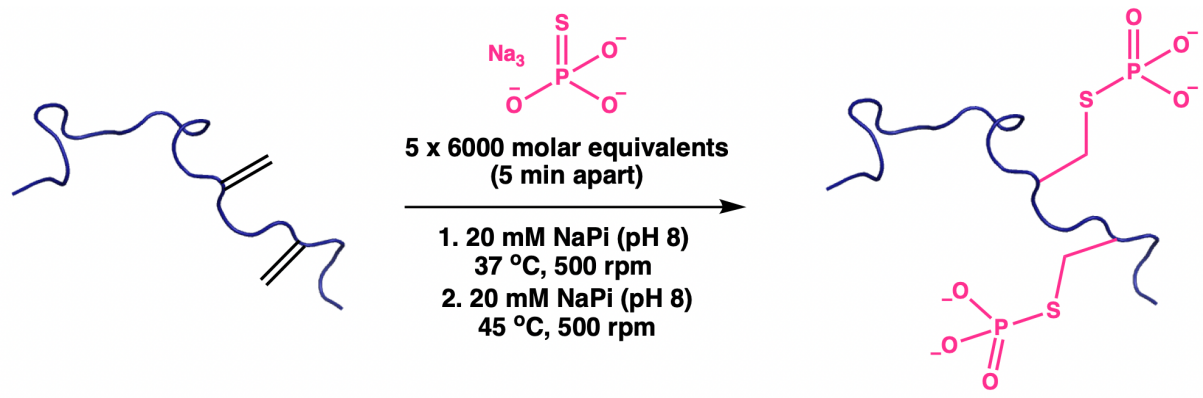
Figure 3.5. pCys installation at single sites. (a) Reaction scheme for pCys installation on single Dha bearing tau mutants. (b) LC-MS spectrum of S199Dha reaction with sodium thiophosphate at 4 hrs, representative of the progress of the reaction at the other positions. (c) LC-MS spectrum of S199pCys after 8 hrs. (d) LC-MS spectrum of S262pCys after 8 hrs. (e) LC-MS spectrum of S356pCys after 8hrs.

3.2.4. Attempted dual phosphorylation reactions at S262Dha & S356Dha mutant

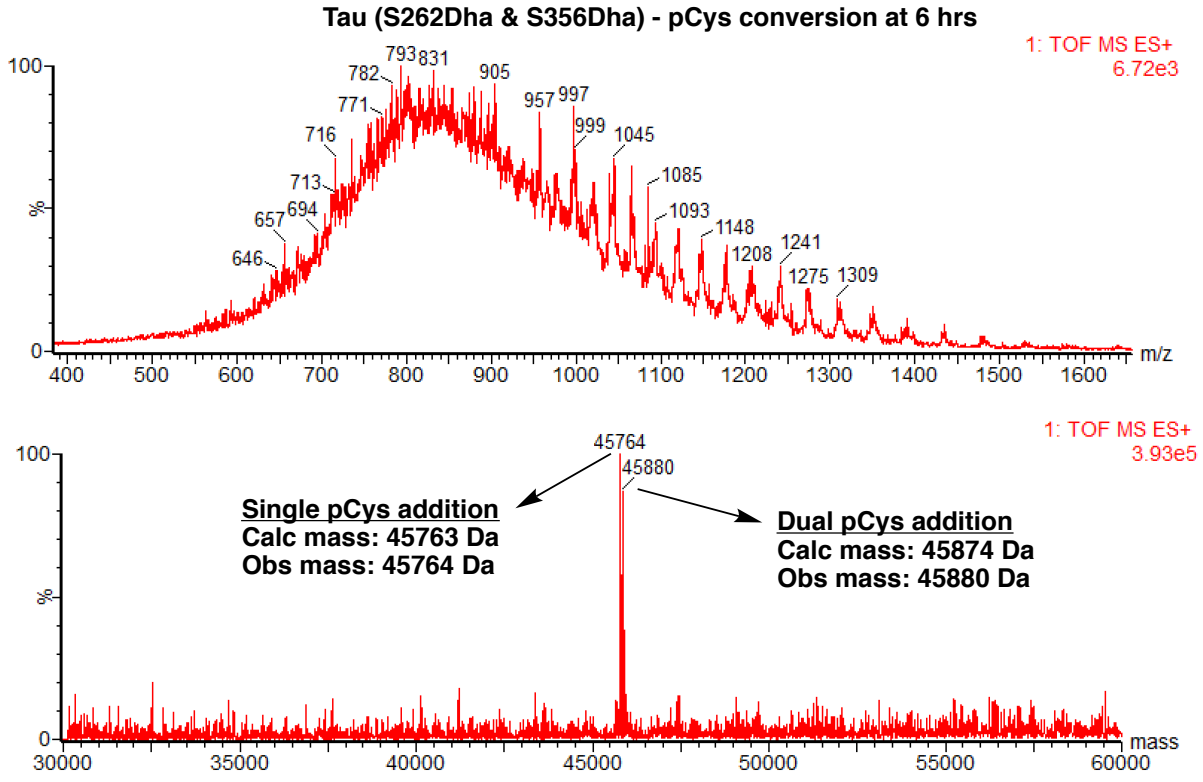
Pursuing dual modifications via Dha is most commonly done on smaller peptides[179], however it has been achieved on full proteins as well. Davis and co-workers were able to install two acetyl lysine mimetics on histone H3 using Dha as an intermediate[160]. While tau would be the largest protein dual modification has been attempted on by far, their succuss gave us hope that tau could be amenable to such an approach. With the S262Dha & S356Dha mutant in hand we initially attempted the dual pCys instalment using the same conditions as the single pCys reactions (**Figure 3.6a**). Checking the reaction after 6 hrs by LC-MS revealed that it had progressed to a near 50/50 mixture of single and dual pCys installation (**Figure 3.6b**). We extended the reaction initially out to 10 hrs and observed that there was about a 75% conversion to the desired dual modification (**Figure 3.6c**). Unfortunately, under these conditions this appeared to be the maximum conversion that we could observe. Extending the reaction out to 24 hrs did not improve conversion of the singly modified population to dual modification and appeared to create some unwanted side products with unexpected mass shifts (**Figure 3.6d**). This was a consistent observation for many attempts to extend the reaction time. We attempted to increase the rate of the reaction by elevating the temperature to 45 ^{o}C (**Figure 3.6a**). To our dismay, checking by LC-MS after 6 hrs did not reveal an enhanced conversion to the desired dual modification (**Figure 3.6e**); on top of that more unexpected mass shifts of undesired side-products were also observed.

A potential explanation for the inability to fully convert tau (S262Dha & S356Dha) to a dual pCys species is that, for a sub-population of the protein, a reactive lysine somewhere in the sequence has performed an intramolecular attack at one of the Dha's. This would be undetectable by LC-MS as the mass of the protein would not change. One solution for this would be to drop the pH of the reaction (down to perhaps ~6) as then nearly all of the lysines on the protein will be protonated and thus incapable of a nucleophilic attack. However, this would also greatly reduce the rate of both the Dha conversions on the protein as well as the subsequent modification by sodium thiosulphate. Unfortunately, due to time constraints and the COVID-19 pandemic limiting lab capacity, we were unable to attempt a reaction using these conditions.

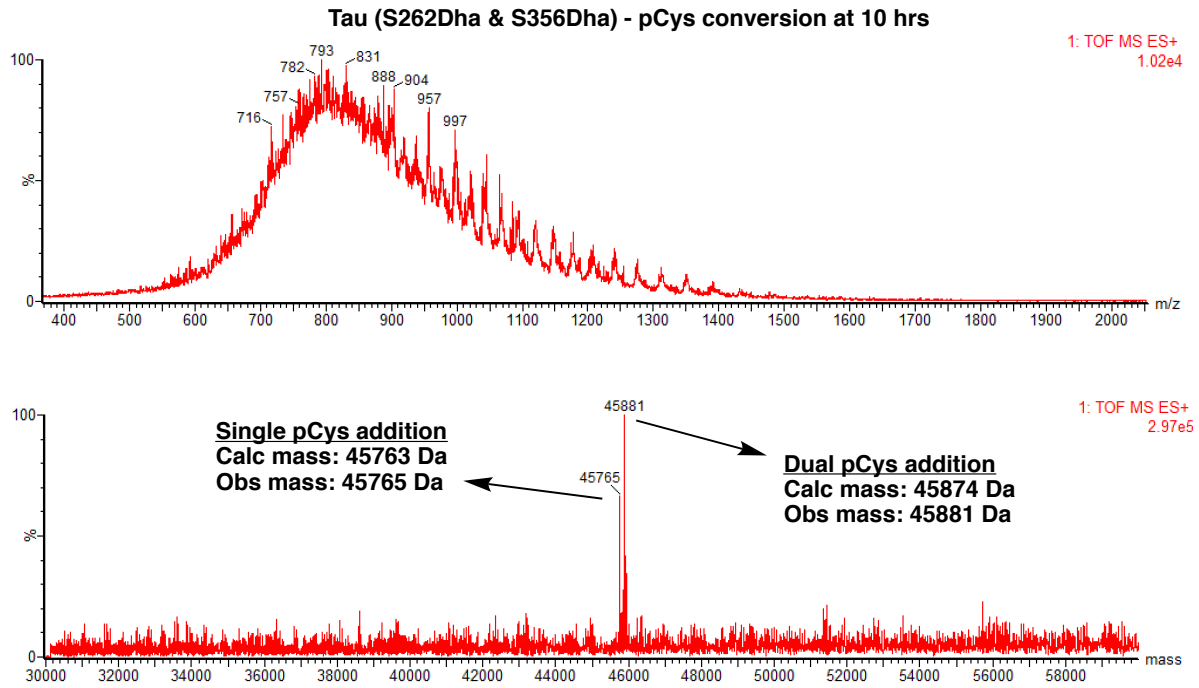
a


Na3
5 x 6000 molar equivalents
(5 min apart)
1. 20 mM NaPi (pH 8)
37 °C, 500 rpm
2. 20 mM NaPi (pH 8)
45 °C, 500 rpm

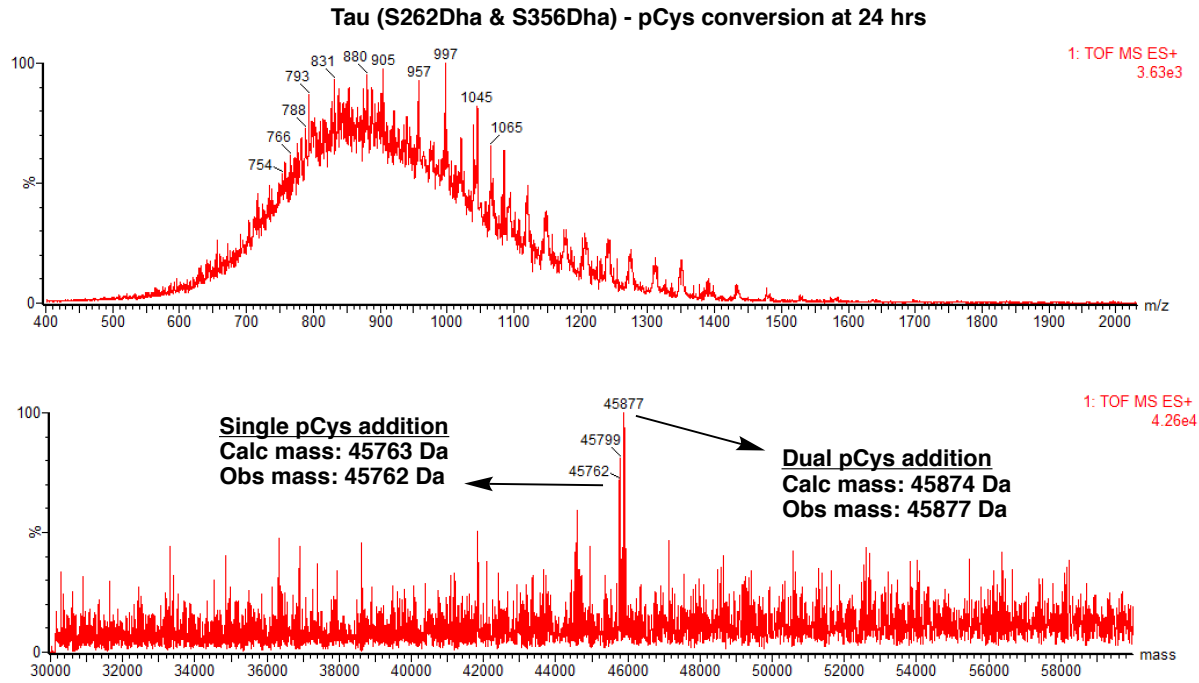
b


Tau (S262Dha & S356Dha) - pCys conversion at 6 hrs
1: TOF MS ES+
6.72e3
100
%
0
646
657
694
713
716
771
782
793
831
905
957
997
999
1045
1085
1093
1148
1208
1241
1275
1309
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
m/z
1: TOF MS ES+
3.93e5
45764
45880
Single pCys addition
Calc mass: 45763 Da
Obs mass: 45764 Da
Dual pCys addition
Calc mass: 45874 Da
Obs mass: 45880 Da
30000
35000
40000
45000
50000
55000
60000
mass

c


Tau (S262Dha & S356Dha) - pCys conversion at 10 hrs
1: TOF MS ES+
1.02e4
793
831
782
888
904
757
957
716
997
m/z
1: TOF MS ES+
2.97e5
45881
Single pCys addition
Calc mass: 45763 Da
Obs mass: 45765 Da
45765
Dual pCys addition
Calc mass: 45874 Da
Obs mass: 45881 Da
mass

d


Tau (S262Dha & S356Dha) - pCys conversion at 24 hrs
1: TOF MS ES+
3.63e3
831
880
905
997
793
957
788
1045
766
1065
754
m/z
1: TOF MS ES+
4.26e4
45877
Single pCys addition
Calc mass: 45763 Da
Obs mass: 45762 Da
45799
45762
Dual pCys addition
Calc mass: 45874 Da
Obs mass: 45877 Da
mass

e

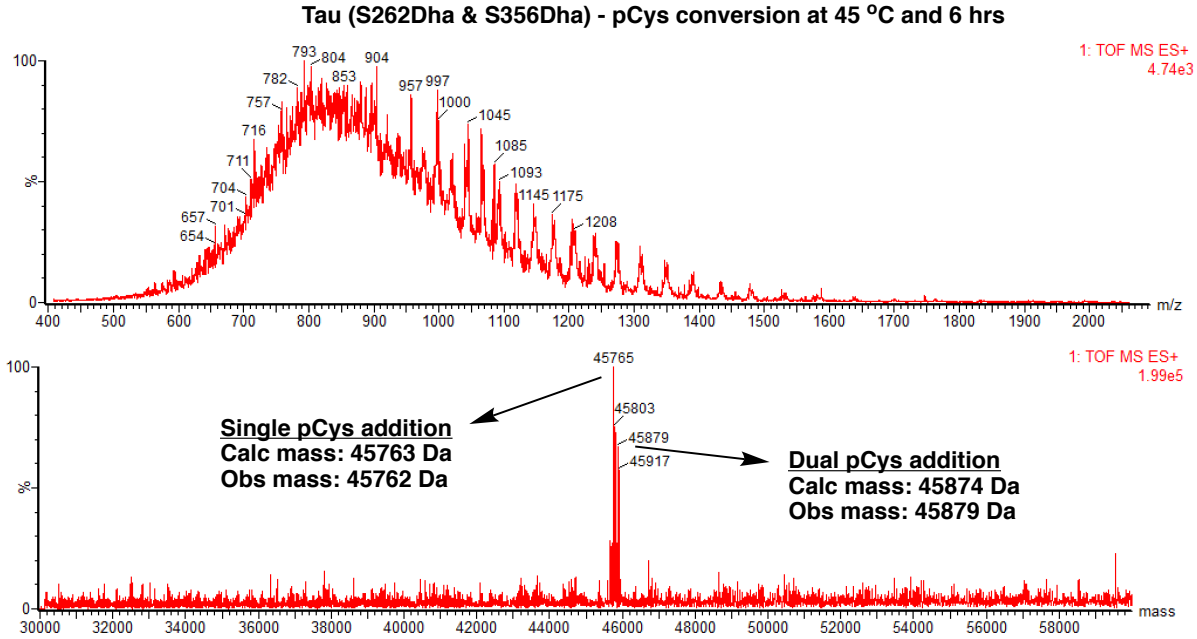
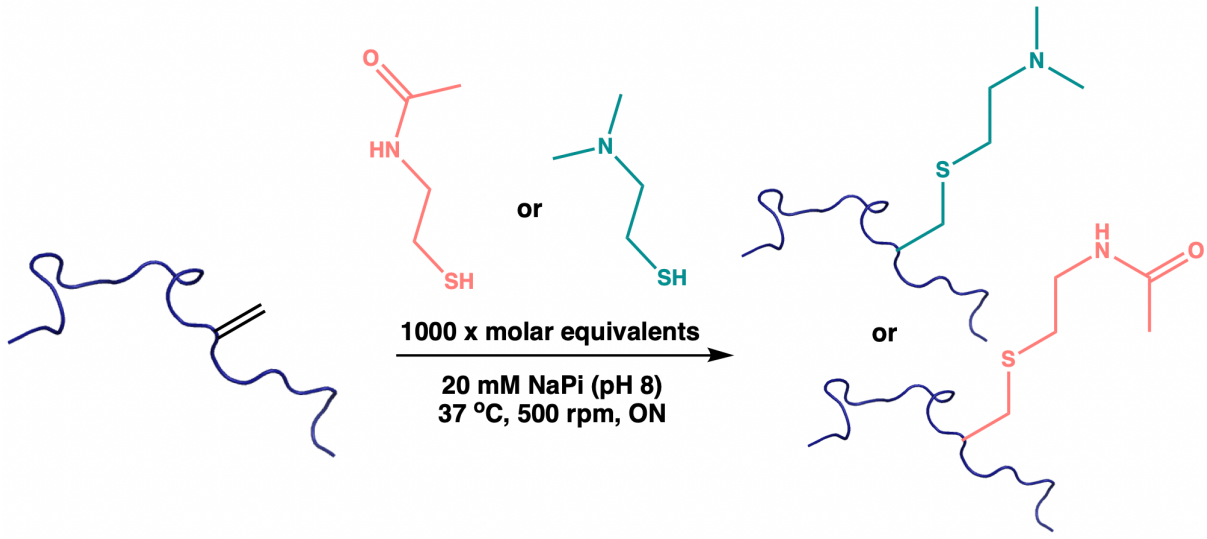


Figure 3.6. Attempted pCys installation on Tau (S262Dha & S356Dha). (a) Reaction schemes for both conditions pursued for this modification. (b) LC-MS of the reaction at 37 °C at 6 hrs. (c) LC-MS of the reaction at 37 °C at 10 hrs. (d) LC-MS of the reaction at 37 °C at 24 hrs (representative of several attempts). (e) LC-MS of the reaction at 45 °C at 6 hrs.

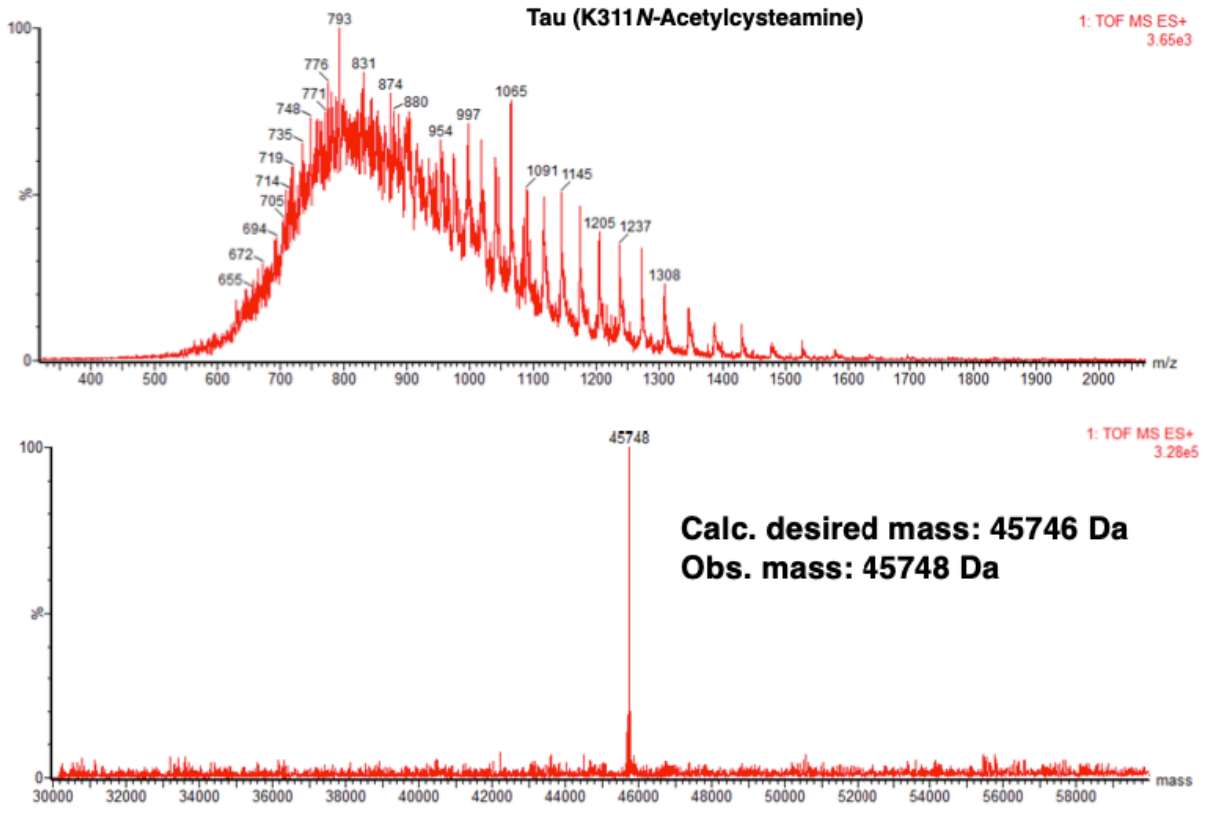
3.2.5. Acetylation and dimethylation mimic installations at K311Dha

While we were pleased with the amenability of Dha bearing tau to single pCys installation, we ideally wanted to prove that chemical mutagenesis can install a diverse array of less commonly studies PTMs on tau. Therefore, we pursued the installation of *N*-acetylcysteamine and captamine at K311Dha. Reaction conditions were the same as previously used for sodium thiophosphate addition (20 mM NaPi [pH 8], 500 rpm) (**Figure 3.7a**). Taking a similar approach to the reactions with the antibodies in chapter 2 we kept the molar excess of these reagents to 1000 equivalents. Allowing the reactions to react overnight afforded the homogenous desired products (**Figure 3.7b,c**)

a


O
HN
SH
or
N
SH
1000 x molar equivalents
20 mM NaPi (pH 8)
37 °C, 500 rpm, ON
N
S
or
H
N
O
S

b


Tau (K311 N-Acetylcysteamine)
1: TOF MS ES+
3.65e3
793
776
831
874
771
880
748
997
1065
735
954
719
714
1091
1145
705
1205
1237
694
672
1308
655
100
%
0
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
m/z
1: TOF MS ES+
3.28e5
45748
Calc. desired mass: 45746 Da
Obs. mass: 45748 Da
100
%
0
30000
32000
34000
36000
38000
40000
42000
44000
46000
48000
50000
52000
54000
56000
58000
mass

c

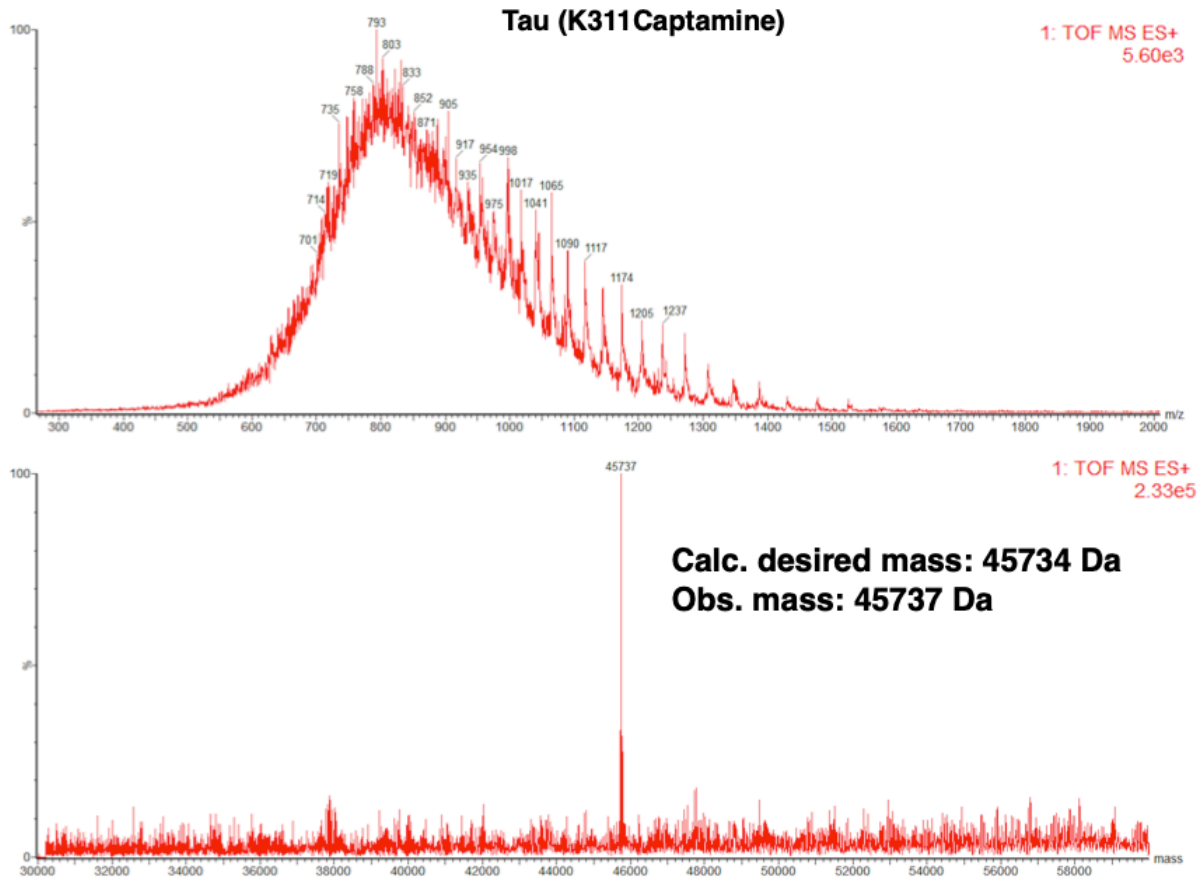


Figure 3.7. PTM mimetic installation at K311Dha. (a) Reaction scheme for reacting tau (K311Dha) with captamine and *N*-acetlycysteamine. (b) LC-MS confirming homogenous addition of *N*-acetlycysteamine. (c) LC-MS confirming homogenous addition of captamine.

3.2.6. Initial functional characterisation of tau chemical mutants

With the site-specific PTMs in hand, we sought to prove their ability to recapitulate known physiological consequences and then to characterise new PTMs in a tubulin polymerisation assay. Before any assays were conducted, the chemical mutants were first inspected by CD to verify no significant structural changes had occurred during the reaction process. Nearly all profiles were in perfect agreement with profile of unmodified tau, the profile for the K311Captamine mutant varied slightly but still had a profile very much matching a disordered structure (Figure 3.8).

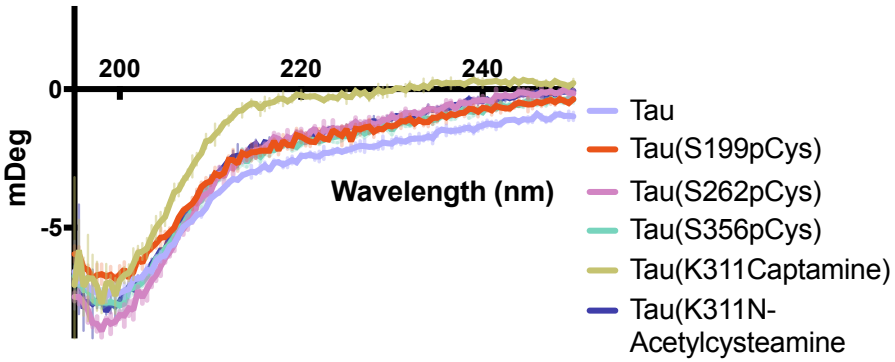


Figure 3.8. Tau chemical mutant CD spectra. CD spectrums of unmodified tau compared to the chemical mutant PTMs.

With the structures largely matching those of unmodified tau, the tubulin polymerisation assay could be pursued. To solutions of 3 mg/mL tubulin in assay buffer (80 mM PIPES [pH 6.9], 2 mM $MgCl_2$, and 0.5 mM EGTA) Tau was added to a final concentration of 15 μM, then the reaction was supplemented with GTP to a final concentration of 1 mM to initiate the polymerisation reaction immediately before inserting the plate and monitoring absorbance at 340 nm. This polymerisation assay is provided by Cytoskeleton Inc. and is a standard in the field of microtubule polymerisation, from studying tau's role to screening small molecule inhibitors of polymerisation. It is also the same kit used by Lashuel and co-workers for all their characterisation of their semi-synthetic tau variants[173,175]. Screening the K311*N*-acetylcysteamine variant first and directly comparing to the data from Ref [173] for the semi-synthetic acetylated K280 revealed that our method exactly recapitulated the activity of this more arduously created construct (**Figure 3.9a,b**).

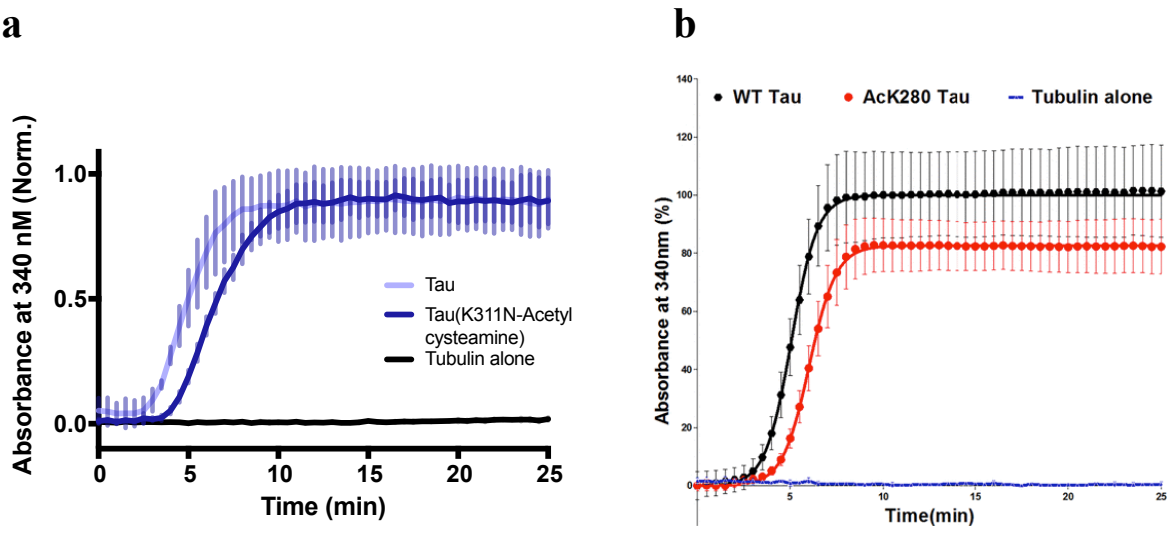
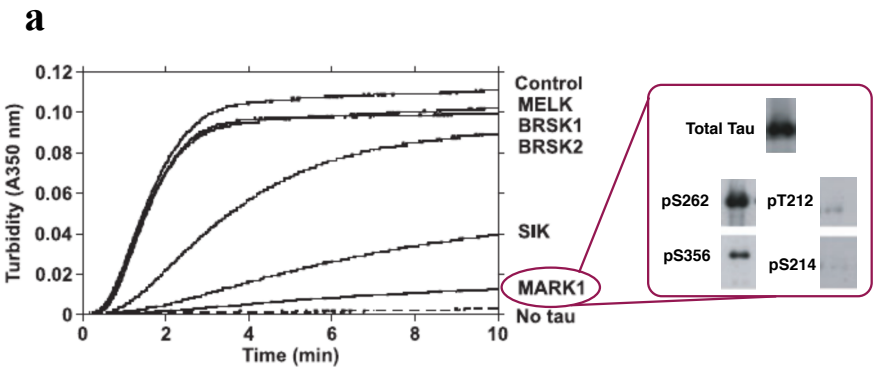


Figure 3.9. Comparison of polymerisation activity of acetylated K311 derived from chemical mutagenesis and acetylated K280 derived from semi-synthesis reveals similar profiles. (a) Tubulin polymerisation profiles of Tau and K311*N*-acetylcysteamine from this study. (b) Polymerisation data for semi-synthesised acetyl-K280 Tau from Ref [173] (adapted with permission).

The semi-synthesis approach has not yet been used to create full length phosphorylated tau at positions 262 and 356. Lashuel and co-workers have only applied this approach to install phosphates at these sites on the K18 fragment (just the four repeated sequences in the MBD) of tau[175]. As this fragment has a very different tubulin polymerisation profile compared to full length Tau, we sought to compare the polymerisation activity of our chemical mutants to *in vitro* phosphorylated tau by the kinase MARK1 that nearly exclusively phosphorylates at S262 and S356 (although to differing extents, and with some phosphorylation at T212 and S214)[171] (**Figure 3.10a**). Our polymerisation reactions with S262pCys and S356pCys revealed that, as expected, both significantly inhibit the rate of tubulin polymerisation, with S356pCys having the greatest inhibitory effect (**Figure 3.10b**). The inhibitory activity we observe is in line with the large decrease in activity observed when tau is phosphorylated *in vitro* with MARK1 (note that the polymerisation conditions in Ref [171] were slightly different, hence the altered time scale). While studies using the phosphomimetic mutations of S262E and S356E are consistent in some regards to our results, such as how the S356E mutation almost completely inhibits polymerisation activity, the activity of S262E displays almost a linear polymerisation behaviour, rather than the delayed but no less sigmoidal profile of our chemical mutant[180]. This discrepancy may serve to highlight the subtleties in activity that are not capable of being recapitulated by cruder mimics.



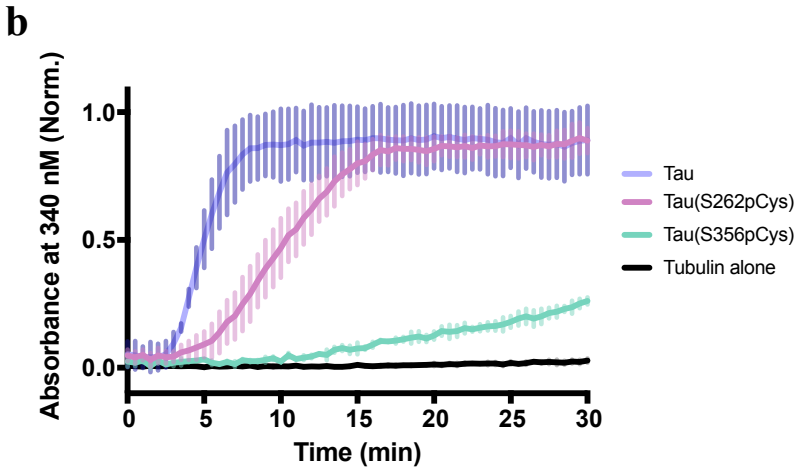


Figure 3.10. Comparison of polymerisation activity of pCys tau mutants at position S262 and S356 to naturally *in vitro* phosphorylated Tau by the MARK1 kinase. (a) Tubulin polymerisation activity and western blots revealing phosphorylation pattern for tau incubated with the kinase MARK1 (adapted with permission from Ref [171]). (b) Tubulin polymerisation activity of the pCys chemical mutants at site 262 and 356 created in this report.

Serendipitously, another member of our group, Dr. Faidon Brotzakis, conducted molecular dynamics simulations of tau in complex with MTs with the aim to predict PTM sites that may regulate MT association. The approach used cryo-EM measurements to determine an ensemble of structures representing the structure and dynamics of a tau-microtubule complex comprising an extended microtubule-binding region of tau (residues 202-395). To predict tau residues that could be important for regulating MT stability, both the average number of contacts through the simulation and the root mean square fluctuations (RMSF) of the residues were quantified. If a residue has a high number of contacts (>7.5) it is likely to be a stabilising interaction and if its RMSF is significant it is also likely accessible for modifications by kinases or acetyltransferases. This analysis rightly predicted the know regulatory phosphorylation sites explored in this report of S262 and S356 (**Figure 3.11a-c**). Strikingly, it also predicted that acetylation at K311 was likely to have a very similar effect as acetylation at K280 that was verified through the results in this report (**Figure 3.11d,e**). Marrying this computational approach with the relatively rapid experimental verification by chemical mutagenesis could be a powerful approach for predicting PTMs regulating protein complex stability when cryo-EM measurements are available, which is becoming increasingly routine to obtain.

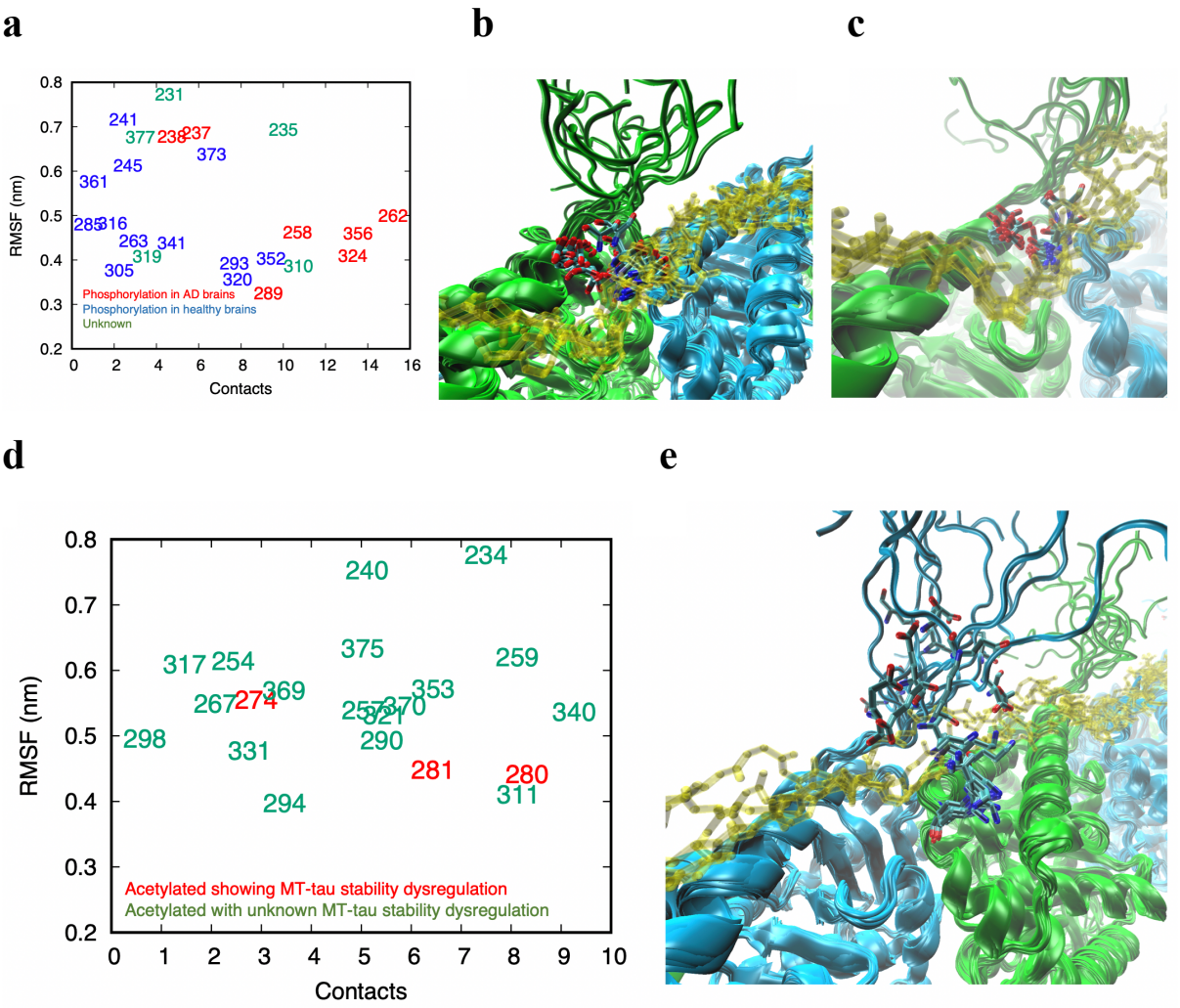


Figure 3.11. Molecular dynamics-based predictions of regulating PTM sites in tau's MBD. (a) RMSF and average contact values of serine residues through the molecular dynamics simulation. (b) Ensemble interactions of S262 of tau with E434 of tubulin. (c) Ensemble interactions of S356 of tau with E434 of tubulin. (d) RMSF and average contact values of lysine residues through the molecular dynamics simulation. (e) Ensemble interactions of K311 of tau with R402 and E436 of tubulin. All figures provided by Dr. Faidon Brotzakis.

With the accuracy of the acetylation and phosphorylation chemical mutant mimetics established, we sought to characterise the polymerisation activity of our poorly understood mimics of K311 dimethylation and S199 phosphorylation. The dimethylation mimic (K311Captamine) showed a very small and likely insignificant delay in tau's polymerisation activity (**Figure 3.11a**). Meanwhile, the S199pCys mutant inhibited tubulin polymerisation to a nearly identical extent as the S262pCys chemical mutant (**Figure 3.11b**). At first glance, this is a very surprising result since S199 is well within the PRD of tau, compared to the very central position of S262 in the MBD. However, several studies in the literature can serve to lend

credence to this result. An early study of the regions of Tau demonstrated that when the PRD is deleted tau is incapable of polymerising tubulin[181]. A more recent study has found that the PRD alone is capable of binding tubulin tightly and stoichiometrically and is capable of initiating polymerisation, and to an even greater extent that the MBD alone[182]. This study goes much further in its characterisation of this domain and shows that the PRD is actually likely the core tubulin binding domain of tau that begins the polymerisation process that is then further enhanced by the MBD, with the entire process being negatively regulated through intramolecular interactions with the N-terminal domain (**Figure 3.11c**). With phosphorylation at S199 a known biomarker for AD progression[176,177], it is possible that this association stems from this modification interfering with the initiation of tubulin polymerisation by the PRD in the model proposed in Ref [182] (**Figure 3.11c**), and thus leading to higher levels of unbound tau capable of aggregating. Although more experiments will be needed to verify this hypothesis.

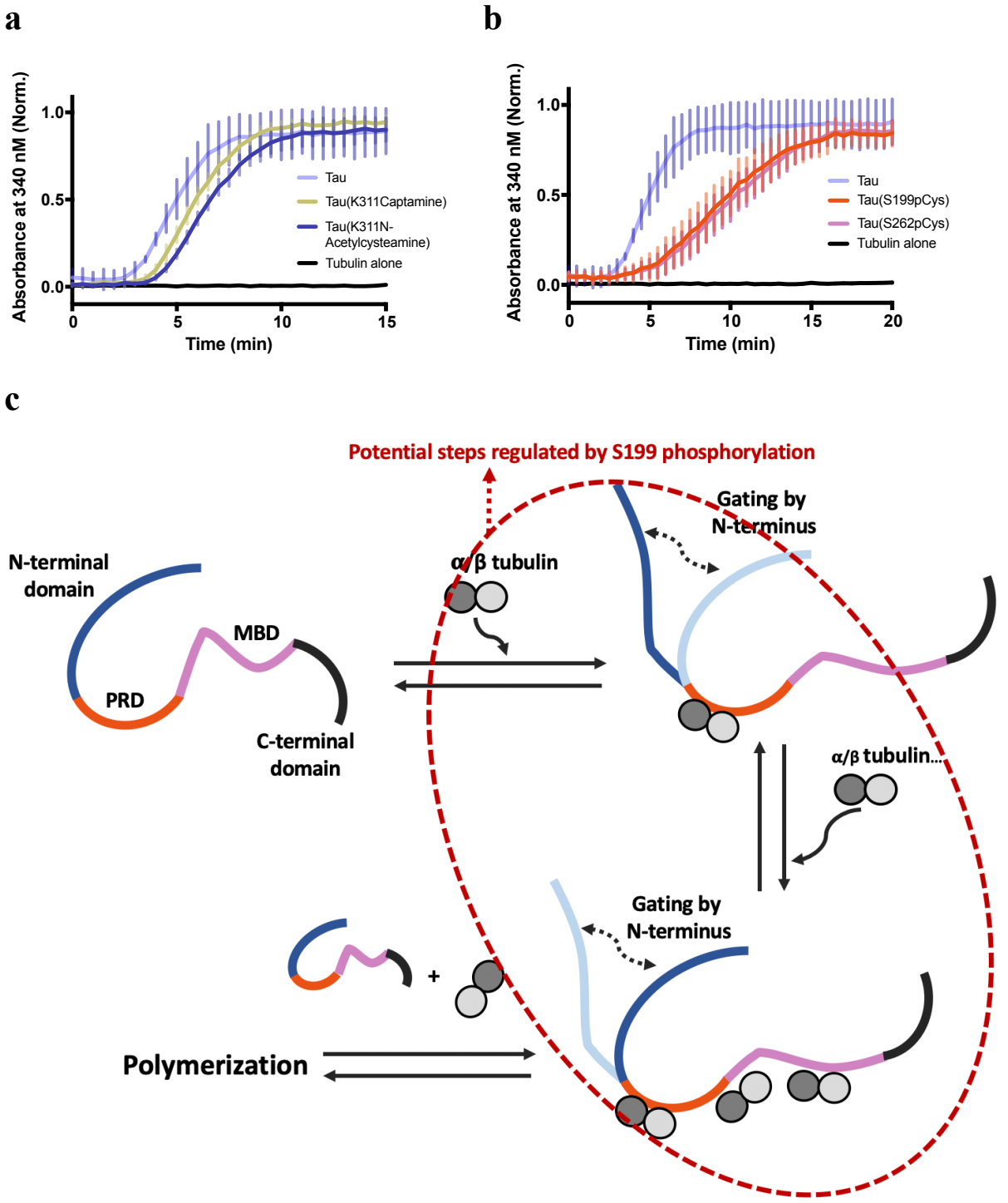


Figure 3.12. Tubulin polymerisation by K311Captamine and S199pCys and polymerisation model. (a) Tubulin polymerisation of Tau (K311Captamine) compared to Tau and Tau (K311*N*-acetylcysteamine). (b) Tubulin polymerisation of Tau (S199pCys) compared to Tau and Tau (S262pCys), showing near perfect overlap of in activity of S199pCys and S262pCys. (c) Initiation of tubulin polymerisation model described in Ref [182] and the potential step that phosphorylation at S199 could disrupt.

3.3 Conclusions

The aim of this chapter was to explore the amenability of Tau for chemical mutagenesis to enable relatively easy access to highly accurate PTMs to facilitate their study. To this end, we pursued the installation of a pCys phosphorylation mimetic via Dha at three different sites, one variant with two pCys modification sites, and two lysine PTM mimetics, acetylation and demethylation, at one site. At the outset, we determined that the commonly used reagent DBHDA for converting cysteine to Dha formed a stable intermediate on the protein and could not be used. However, we found that a recently developed more reactive molecule for this purpose, MDBP, effectively installed Dha at a reasonable timescale for all sites investigated in this report. The subsequent additions of the PTM mimetics went smoothly for all of the tau variants containing single Dha residues. Dual modification at the S262Dha & S356Dha double mutant proved to be difficult to reach homogenous conversion to the dual pCys mutation, possibly due to an intramolecular attack by a reactive lysine. Unfortunately, due to time-constraints and the limited work capacity in the lab during the COVID-19 pandemic this could not be optimise further.

With the successful production of the single chemical mutants we performed an initial characterisation of their effects on tau's important physiological function of tubulin polymerisation. The data from the polymerisation of our K311 acetylation mimic greatly supported the physiological validity of our chemical mutants as it perfectly replicated the polymerisation profile of acetylated K280 (located within a similar motif) generated through a laborious semi-synthesis protocol. This was then further corroborated with the polymerisation data for S262pCys and S356pCys, which both showed a significant inhibition of activity in line with data from *in vitro* phosphorylated tau by the kinase MARK1 that phosphorylates primarily at these sites. Validating the accuracy of these chemically installed PTMs was a very important goal of this project, as we can now have a faith in the results from newly explored PTMs. Of the two new PTMs investigated in this early characterisation, we observed an initially unexpected significant inhibition of polymerisation by the installation of pCys at S199 in the PRD. However, looking at both distant and recent literature on the PRD and its role in tau function provides a rational explanation as the PRD has an increasingly supported role in the initiation and elongation steps of tubulin polymerisation that has largely been underappreciated[182]. While more work will need to be carried out to fully validate this hypothesis, from our data it appears that phosphorylation at S199 may play a role in regulating

the capability of the PRD to associate with tubulin and may form a basis for explaining this modification's pathological association. This tantalizing result serves to highlight the potential for chemical mutagenesis to facilitate the study of tau PTMs and their undoubtably numerous functional consequences. We hope that this brief and primarily methodological report will enable the field with a relatively facile method, and we look forward to further results on the potential role of S199 phosphorylation as well.

Chapter 4: Discussion and future work

4.1 Chapter 2 discussion and future work

Since the beginning of the research reported in chapter 2, a plethora of studies have been published with the aim of augmenting protein activity using non-canonical amino acids. New catalytic mechanisms, previously impossible for natural enzymes, have been enabled through the incorporation of non-canonical residues in active sites[183,184]. While these studies relied on genetic code-expansion technology, they still highlight the growing use case for expanding side-chain diversity. In the context of protein-protein recognition, Rogers and co-workers provided a seminal study validating the utility of non-canonical amino acids in augmenting binding interactions[185]. Using an *in vitro* translation system with flexizymes (flexible tRNA-acylation ribozymes that can load many non-canonical amino acids) the researchers were able to incorporate 21 non-canonical amino acids throughout the 27-amino acid sequence of the BH3 domain of the protein PUMA. The binding of this domain to its partner, the protein MCL1, has been a model system for studying protein-protein interactions[186]. Using this approach, they were able to find and combine 5 potentiating mutations that increased the affinity 25-fold over the natural sequence (from an already potent K_D of 4 nM to 160 pM). In line with our findings, many of the single substitutions with non-canonical side chains led to a nearly complete loss of activity in this system. They also observed that changes as small as single methyl groups led to drastic changes in binding energies. Together with our results, it appears that the "magic methyl" effect long seen in small molecule drug design may extend to the case of non-canonical amino acids in protein-protein interactions[187].

Along with these studies supporting the general use of non-canonical sidechains, the chemistry available for chemical mutagenesis has also been advanced since the beginning of the work in this report, both for the installation and modification of Dha. Yang and co-workers developed a proximity enabled intramolecular reaction to install Dha or its close analogue dehydrobutyrine (Dhb, addition of a single methyl group)[188]. The installation relies on the incorporation of the UAA fluorosulfate-L-tyrosine through genetic code expansion proximally to a Ser or Thr residue where Dha/Dhb is ultimately desired; at the end of the reaction the fluorosulfate-L-tyrosine is converted to normal tyrosine. This system allows for the incorporation of Dha/Dhb into proteins that may not be amenable to extended reaction

conditions with traditional reagents and can even install Dha/Dhb in living systems. Along with this, a new and improved reaction for the installation of $C(sp^3)$-$C(sp^3)$ bonds at Dha has also been recently reported[189]. The Davis lab again leads the way in this innovation. Instead of using halogen bearing chemical mutant precursors, this new approach relies on boronic acid catechol ester precursors that can generate $RH_2C\bullet$ radicals when illuminated with blue light (450 nm) in the presence of a widely available ruthenium catalyst in a range of pH's and temperatures. These chemical advances should further the amenability of different substrates to chemical mutagenesis mediated side-chain diversification.

Our results for using chemical mutagenesis to systematically enhance antibody activity will hopefully motivate more research into this application of chemical mutagenesis. Due to the nature of the target of the starting antibody in this study, we had to take an empirical approach initially and screen a large variety of sidechains. In the future, an interesting application would be to use SAR via chemical mutagenesis to rationally improve the activity of an antibody using structural information of it bound to the antigen. Iterative rounds of side-chain diversification followed by activity screening and structural characterisation can then truly follow the traditional workflow taken by medicinal chemists. This may prove to be an appropriate route to activity maturation for starting candidates targeting highly conserved clinically relevant epitopes, such as those of viral surface proteins. Especially when the potential development of escape mutations in most protein regions is very high[190]. Another very recent and exciting development in protein-based therapeutics that could be amenable to this chemical mutagenesis approach is that of proximity enabled reactive therapeutics (PERx)[191]. The Wang group at UCSF has just developed the first example of such a therapeutic. This molecule interferes in the programmed cell death protein-1 (PD-1) signalling pathway, which is a key mediator of suppressing the immune response in cancers and many other diseases[192]. In their approach, a latently reactive UAA was installed at a critical site within recombinant programmed cell death protein-1 (PD-1). This then covalently reacts upon binding to its ligand (PD-L1) on the surface of cancer cells and prevents them from evading destruction by capping this signalling domain with the inert PD-1. Chemical mutagenesis on antibodies could prove to be an excellent means to develop similar therapeutics for a variety of targets. The ability to install a wide variety of potential PERx appropriate sidechains at many sites through a post-expression protocol would greatly accelerate the optimisation process, rather than expressing/purifying many variants with UAAs introduced through genetic codon expansion technology. A potential use case for applying PERx with antibodies in AD would be for an antibody that specifically targets the secondary nucleation sites along the fibril

surface. When such an antibody is developed, if it is merged with PERx through chemical mutagenesis, it would in effect nullify the production of toxic oligomers permanently. As evident through these developments in the field during the work carried out in this thesis, the application of advanced chemical biology tools in biotechnology is happening at an ever-increasing pace. We anticipate that the application of chemical mutagenesis for antibody development will exceed that of SAR studies alone in the future.

4.2 Chapter 3 discussion and future work

Large advances in the use of chemical mutagenesis to study of ever more complex PTMs has also occurred since the beginning of the work on Tau reported in this thesis. Perhaps the most striking example of this has been the elucidation of the multiphase reaction mechanism of glycogenin in initiation of glycogenesis[174]. This study did not utilise Dha, but instead a very clever incorporation of another synthetically versatile non-canonical amino acid *p*-iodophenylalanine in place of a catalytic tyrosine residue (Tyr195). In the cell, this tyrosine acts as the initiator site of the polymerisation of glycogen through a stepwise autocatalytic process. By replacing this residue with *p*-iodophenylalanine, the researchers were able to precisely probe activity by conjugating boronic acid bearing oligosaccharides of specific lengths and compositions to this site through a palladium mediated Suzuki-Miyaura cross coupling reaction. Through the installation of these precise intermediates, the researchers revealed a triphasic kinetic mechanism for the autoglucosylation reaction and a "proof-reading" mechanism of the enzyme to ensure the fidelity of the growing polysaccharide. This study is an elegant example of using a different reaction available to chemical biologists to study increasingly complex PTM mediated protein behaviour.

The significance of diverse tau PTMs on its pathogenesis has only become more apparent since the beginning of this work. The Fitzpatrick lab recently made the first definitive observation of PTMs influencing the formation of strain-specific structures in individual tauopathies[193]. Using patient derived amyloid fibrils from AD and corticobasal degeneration (CBD), the team combined cryo-EM structural determination with mass-spectrometry to map PTM sites and their influence on the distinct fibril morphologies in these two tauopathies. This comprehensive biophysical characterisation revealed that CBD fibrils are heavily post-translationally modified compared to AD derived fibrils, and that specific sites with consistent ubiquitination in CBD were key mediators of inter-protofilament interfaces, heavily

influencing the morphology of the fibrils. Their analysis also elucidated that acetylation at K311 was a key modification in AD fibrils that determined the favourability of forming paired helical filaments over straight filaments. More generally, they also observed that lysine acetylation likely creates favourable conditions for β-strand stacking within the fibrils.

Sadly, the plans for our studies of tau PTMs was cut short because of restrictions put in place due to the ongoing pandemic. However, the preliminary data presented here on the amenability of tau to chemical mutagenesis and the physiological accuracy of the PTM mimetics is very encouraging for the application of this technique to study tau modifications. One of the most intriguing results was the effect of phosphorylation at Ser 199 on tubulin polymerisation. Future work is needed to definitively establish if this modification is truly involved in the regulation of the PRD's role in tubulin association. This could be probed by purifying tau truncated to solely the PRD and accurately quantifying the effects of the S199pCys modification on tubulin binding. An initial characterisation could be carried out using BLI in order to get detailed kinetic data on the PTM's effect. Elucidating all the mechanisms behind S199 phosphorylation's pathological association will also likely need some characterisation of its effect on tau's fibrillization. Unfortunately, a physiologically relevant *in vitro* aggregation assay has not yet been developed for full-length tau, and aggregation studies using fragments almost exclusively focus on variations of the MBD and C-terminal region. Therefore, to fully explore the potential effects of this modification, aggregation screens would likely need to be conducted with a series of truncated tau variants containing the PRD and varying regions of the MDB and C-terminus.

To more comprehensively study PTMs on tau and their potential ability to "cross-talk," the dual modification reaction will also need to be optimised in the future. The working hypothesis of an intramolecular attack by a reactive lysine as the reason behind the inability of the S262Dha & S356Dha to react to completion with sodium thiophosphate could be assessed by LC-MS/MS. If this is indeed what is occurring, then the pH of the reaction will need to be dropped in order to nullify the lysine residues. This could also be an issue that is unique to these positions if this is the case. Once this hurdle is overcome, studying the cumulative effects of PTMs on tau behaviour will be more readily accessible than the established semi-synthesis method. When these protocols for chemical mutagenesis on tau are published, we hope this will provide a widely accessible method for the field to accurately study the effects of PTMs on this crucial protein.

4.3 Potential of chemical mutagenesis

The overarching aim of the work described in this thesis was to explore a new application of chemical mutagenesis and to apply established techniques within the field to a new and important protein, all within the context of Alzheimer's disease. These goals fit into the greater context within chemical biology of widening the utility and use cases of chemical tools. While nature is largely limited to synthesising proteins with the 20 canonical amino acids, life greatly expands the functions and behaviour of proteins through the myriad of cellular systems responsible for PTMs. Chemically simple PTMs, such as methylation, to more complex modifications, such as glycosylation, have profound influences on biological processes in all kingdoms of life. To enable true synthetic biology, life scientists will need powerful techniques to rationally expand the chemical space of proteins. Due to the inherent limitations of the protein biosynthesis machinery, approaches beyond gene-codon expansion technology must be embraced to accomplish this. Ideally, these methods will not be overly complex and could be rapidly adopted by the field as a whole, similar to how site-directed mutagenesis is implemented in labs across the world today. This will enable the development of synthetic biology in the coming century to match the advances seen in synthetic chemistry during the previous century. The applications explored in the research presented here of enabling a SAR study to augment protein function and to investigate disease relevant PTMs have wide applicability to many areas of biology and biotechnology. With the accelerating pace that chemical mutagenesis is being readopted and improved, there are likely many more applications to come that we cannot yet envisage. This is a truly exciting time to be at the interface of chemistry and biology and witness their ever more interwoven advances.

Chapter 5: Materials and methods

5.1 DesAB-Aß $_{3-9}$ and its mutants' expression and purification

All antibodies were expressed with an N-terminal hexa-histidine tag from a pRSET-B vector in *E. coli* BL21 (DE3)-Gold strain (Agilent Technologies). Cultures were grown in modified Terrific Broth media (Sigma-Aldrich) supplemented with ampicillin (100 μg/ml) at 37 ºC and induced with 1 mM isopropyl β- d-1-thiogalactopyranoside (IPTG) at an optical density (OD) of approximately 0.8 and were then left to express overnight at 28 ºC. Cells were harvested by centrifugation and re-suspended in standard phosphate buffered saline (PBS) with the addition of one EDTA-Free Complete Protease Inhibitor Cocktail Tablet (Roche) per 500 mL of cell culture and subsequently lysed by sonication (5 min total, 15 s on and 45 s off, 40% amp, on ice). Cellular debris was removed by centrifugation at 15,000 rpm (JA-20 rotor, Beckman Coulter). The cleared lysate was then loaded onto loose Ni^{2+}-NTA resin (Thermo Fisher) previously equilibrated with PBS containing 10 mM imidazole. After washing the loaded resin with PBS containing 40 mM imidazole the protein was eluted with buffer containing 200 mM imidazole. The excess imidazole was subsequently removed during size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 pg column (GE Healthcare LifeSciences, Little Chalfont, U.K.) into PBS, previous to the SEC the protein was incubated with 1 mM dithiothreitol (DTT) for 30 minutes to reduce any dimers that may have formed. The sequence of the singe-domain antibody used is as follows, with the designed CDR3 loop highlighted in bold and the thermally stabilizing mutation from Ref [116] is in red:
N-terminus-
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPKLEVQLVESGGGLVQPGGSL
RLSCAASGFNIKDTYIGWVRRAPGKGKEWVASIYPTNGYTRYADSVKGRFTISADTS
KNTAYLQMNSLRAEDTAVYYCAAGS**HETLTLR**EEEAAAWGQGTLVTVSSGT
-C- terminus

5.2 Tau expression and purification

Tau and its cysteine mutants used in this study were expressed from BL21 Gold DE3 cells (Agilent). Batches of 6 x 1 L cultures of LB medium with 100 μg/mL ampicillin were

inoculated with 5 mL of overnight culture and grown at 37 ºC until an OD 600 nm reached 0.6. Expression was then induced with the addition of IPTG to a final concentration of 0.4 mM, whereupon the temperature was reduced to 18 ºC and the cultures were allowed to express overnight. Cells were harvested in the morning by centrifugation at 6000 g for 30 min at 4 ºC. Pelleted cells were resuspended in lysis buffer (50 mM MES [pH 6.5], 5 mM DTT, 0.1 mM PMSF) and lysed on ice via sonication (1.5 min total, 5 s on/10 s off, 40% amplitude). Lysate was cleared via centrifugation at 18000g for 30 min at 4ºC. To the supernatant DNAse and RNAse was added (0.1 mg/mL) and allowed to sit at room temperature for 5 minutes. Supernatant was then passed through a 0.45 µm syringe filter to remove any particulates. Cation exchange chromatography was then performed using a Hitrap CaptoS column (GE Healthcare) equilibrated with 5 CV ddH_2O, then 5 CV lysis buffer + 1 M NaCl, then 5 CV lysis buffer. The lysate was then loaded at 5 mL/min and washed with 10 CV of lysis buffer. The protein was eluted at a rate of 5 mL/min with 5 mL fractions being collected through a 0-500 mM gradient of NaCl over 10 CV. Fractions were assessed via SDS-PAGE for the presence of Tau and all appropriate fractions were pooled and precipitated with ammonium sulphate (20% w/v) and left overnight on ice. Precipitated proteins were pelleted at 16000 g for 30 min at 4 ºC and then resuspended in 1 x SSPE buffer with 5 mM DTT. Size exclusion chromatography was then performed using a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated in the same resuspension buffer. Proteins were injected and ran at a flow rate of 0.75 mL/min. Fractions corresponding to the appropriate UV signal were pooled and assessed by SDS-PAGE for purity, only the purest fractions of Tau were kept for experiments and the rest was pooled and frozen for subsequent SECs to isolate more pure Tau when needed.

5.3 Site-directed mutagenesis

All site-directed mutagenesis carried out in this thesis was done using the methods described in Ref [169]. Mutagenic primers were designed with overlapping sequences at the 5' end and non-overlapping portions at the 3' end of the sequence. This allows for the primers to use their PCR products as templates as the non-overlapping regions can bridge the "nicks" in DNA of newly synthesised products. Non-overlapping regions were designed to have a T_m 5 to 10 ºC higher than the overlapping regions of the primers, which contained the desired mutations. All primers were ordered from Sigma Aldrich. PCR reactions were conducted in a total volume or 50 µL with reagents supplied from New England Biolabs (NEB) (10 µL 5x

reaction buffer, 250 µM dNTPs, 1 µg/mL each primer, 20 ng starting template, 2 µL DMSO, 0.5 µL Phusion HF polymerase). The thermocycler conditions were as follows: 30 s initial denaturation (98 ºC), 10 s denaturation (98 ºC), 30 s annealing at T_m overlap -5 ºC, 5 minutes elongation (72 ºC) with the previous 3 steps repeated for 30 cycles, a last annealing step was performed for 30 s at T_m overlap -5 ºC and then the last extension was for 7 minutes. PCR reactions were then treated with fast digest DPN1 (1 µL, NEB) at 37 ºC for 4 hrs to digest any starting template. Samples were then cleaned using a Qiagen PCR clean kit before being transformed into XL-10 Blue ultracompetent cells (Agilent). Colonies from the transformations had the plasmids harvested for sequencing using a Qiagen miniprep kit after growing in 5 mL LB cultures overnight. All primers used in this report can be seen below:

H136C-

Fwd: 5'- *GGATCTTGCGAAACC*CTGACCCTGCGCGAGGA-3'

Rev: 5'- *GGTTTCGCAAGATCC*CGCTGCGCAATAATACACAGC-3'

T_m Overlap: 50.6 C

T_m Non-overlap: 59.8 C Fwd, 58.4 C Rev

E137C-

Fwd: 5'- *ATCTCATTGCACCCTGA*CCCTGCGCGAGGAAGA-3'

Rev: 5'- *TCAGGGTGCAATGAGAT*CCCGCTGCGCAATAATAC-3'

T_m Overlap: 49.9 C

T_m Non-overlap: 55.9 C Fwd, 56.3 C Rev

T138C-

Fwd: 5'-*CATGAATGCCTGACCC*TGCGCGAGGAAGAGGCG-3'

Rev: 5'-*GGGTCAGGCATTCATG*AGATCCCGCTGCGCAATAATA-3'

T_m Overlap: 50.6 C

T_m Non-overlap: 59.8 C Fwd, 59.5 C Rev

L139C-

Fwd: 5'- *TGAAACCTGCACCCTG*CGCGAGGAAGAGGCGG-3'

Rev: 5'-*CAGGGTGCAGGTTTCA*TGAGATCCCGCTGCGCAA-3'

T_m Overlap: 50.6 C

T_m Non-overlap: 58.4 C Fwd, 58.4 C Rev

T140C-

Fwd: 5'-*ACCCTGTGCCTGCG*CGAGGAAGAGGCGGCC-3'

Rev: 5'-*CGCAGGCACAGGGT*TTCATGAGATCCCGCTGCG-5'

T_m Overlap: 49.4 C

T_m Non-overlap: 58.4C Fwd, 59.5C Rev

L141C-

Fwd: 5'-*CTGACCTGCCGCGA*GGAAGAGGCGGCCGC-3'

Rev: 5'-*TCGCGGCAGGTCAG*GGTTTCATGAGATCCCGCT-3'

T_m Overlap: 49.4 C

T_m Non-overlap: 57 C Fwd, 57.5 C Rev

R142C-

Fwd: 5'-*ACCCTGTGCGAGGAA*GAGGCGGCCGCGTG-3'

Rev: 5'-*TTCCTCGCACAGGGT*CAGGGTTTCATGAGATCCC-3'

T_m Overlap: 48.8 C

T_m Non-overlap: 55.6 C Fwd, 56.3 C Rev

T138K-

Rev: 5'-*GGGTCAGCTTTTCATG*AGATCCCGCTGCGCAATAATA-3'

Fwd: 5'- *CATGAAAAGCTGACCC*TGCGCGAGGAAGAGGCG-3'

T_m Overlap= 48.2 C

T_m Non-overlap = 59.6 C

E80K-

Fwd: 5'-*GAAAGGCAAGGAATGGG*TGGCGAGCATTTATCCGACC-3'

Rev: 5'-*CCCATTCCTTGCCTTTC*CCGGGTGCACGACGCA-3'

T_m Overlap = 52.4 C

T_m Non-overlap = 60.5 C Fwd, 58.4 C Rev

Tau S262C-

Fwd: 5'- *CGGTTGTACCGAAAACC*TGAAACACCAGCCGGGTGG-3'

Rev: 5'- *GGTTTTCGGTACAACCG*ATTTTAGATTTAACGTTTTTCAGGTCC-3'

T_m Overlap = 52.4 C

T_m Non-overlap = 61.6 C Fwd, 60.8 C Rev

Tau S356C-

Fwd: 5'- *CGGTTGTCTGGACAAC*ATCACCCACGTTCCGGGTG-3'

Rev: 5'- *GTTGTCCAGACAACCG*ATTTTAGACTGAACACGGTCTTTG-3'

T_m Overlap = 52.4 C

T_m Non-overlap = 61.6 C Fwd, 60.3 C Rev

Tau K311C-

Fwd: 5'-*TCGTTTACTGCCCGGTTG*ACCTGTCTAAAGTTACCTCTAAATCTG-3'

Rev: 5'-*CAACCGGGCAGTAAACGA*TCTGAACAGAACCACCACCCG-3'

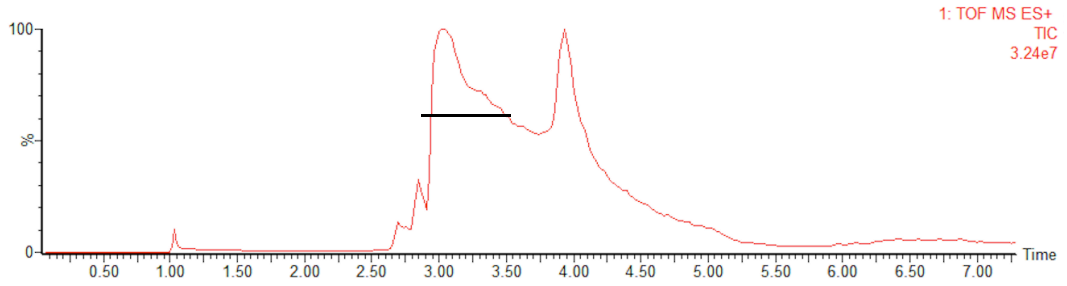
T_m Overlap = 56.3 C

T_m Non-overlap Fwd = 63.7 C Fwd, 63.2 C Rev

5.4 Protein liquid chromatography - mass spectrometry (LC-MS)

Protein LC–MS was performed on a Xevo G2-S TOF mass spectrometer coupled to an Acquity UPLC system using an Acquity UPLC BEH300 C4 column (1.7 µm, 2.1 mm × 50 mm). Water with 0.1% formic acid (solvent A) and 95% MeCN and 5% water with 0.1% formic acid (solvent B) were used as the mobile phase at a flow rate of 0.2 mL/min. The gradient was programmed as follows: 95% A for 0.93 min, then a gradient to 100% B over 4.28 min, then 100% B for 1.04 minutes, then a gradient to 95% A over 1.04 min. The electrospray source was operated with a capillary voltage of 2.0 kV and a cone voltage of 40 V. Nitrogen was used as the desolvation gas at a total flow of 850 L/h. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (v4.1 from Waters) according to the manufacturer's instructions. Representative total ion chromatographs (TICs) for the antibodies and Tau and the regions integrated for analysis are shown below (**Figure 5.1a,b**).

a



b

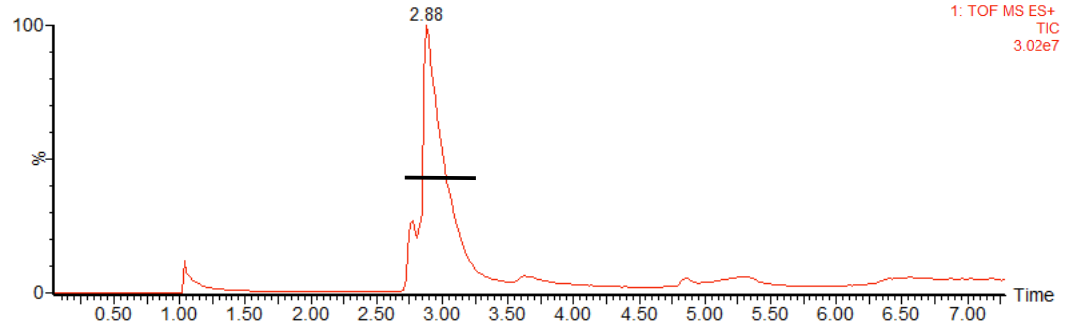


Figure 5.1. Representative TICs for the antibodies and for Tau. (a) Representative TIC from DesAB-A$\beta_{3\text{-}9}$(T138C). (b) Representative TIC of Tau from Tau(S356C). Black bars indicate the regions of the chromatograms containing the desired protein ions.

5.5 Liquid chromatography with tandem mass spectrometry (LC-MS/MS)

Protein solutions were reduced with DTT, alkylated with iodoacetamide, and subjected to enzymatic digestion with chymotrypsin at 37 °C. After digestion, the peptide solutions were pipetted into sample vials and loaded onto an autosampler for automated LC-MS/MS analysis. All LC-MS/MS experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC (Thermo Fisher Scientific Inc, Waltham, MA, USA) system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). The separation of peptides was performed by reverse-phase chromatography at a flow rate of 300 nL/min and a Thermo Scientific reverse-phase nano Easy-spray column (Thermo Scientific PepMap C18, 2 µm particle size, 100 Å pore size, 75 µm inside diameter (i.d.) x 50 cm length). Peptides were loaded onto a pre-column (Thermo Scientific PepMap 100 C18, 5 µm particle size, 100 Å pore size, 300 µm i.d. x 5 mm length) from the Ultimate 3000 autosampler with 0.1% formic acid for 3 minutes at a flow rate of 10 µL/min. After this period, the column valve was switched to allow elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was 80% acetonitrile, 20% water + 0.1% formic acid. The linear gradient employed was 2-40% B in 30 minutes. The LC eluant was sprayed into the mass spectrometer by means of an Easy-Spray source (Thermo Fisher Scientific Inc.). All m/z values of eluting ions were measured in an Orbitrap mass analyzer, set at a resolution of 70,000 and was scanned between m/z 380-1500. Data dependent scans (top 20) were employed to automatically isolate and generate fragment ions by higher energy collisional dissociation (HCD) and 25% normalized collisional energy (NCE) in the HCD collision cell and measurement of the resulting fragment ions was performed in the Orbitrap analyzer, set at a resolution of 17,500. Singly charged ions and ions with unassigned charge states were excluded from being selected for MS/MS and a dynamic exclusion window of 20 seconds was employed. All MS/MS data were converted to mgf files and the files were then submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against a custom database containing the UniProt human database and four sequences relating the 4 proteins which were analyzed. Variable modifications of oxidation (M), deamidation (NQ) and beta

mercaptoethanol and a fixed modification of carbamidomethyl were applied. The peptide and fragment mass tolerances were set to 20 ppm and 0.1 Da, respectively. A significance threshold value of $p<0.05$ and a peptide cut-off score of 20 were also applied.

5.6 Synthesis of 2-mercaptoethyl-N,N,N-trimethylammonium chloride

5 g of 2-(bromoethyl)-trimethylammonium bromide (1 eq, 4.88 mmol) was transferred into a 250 mL round bottom flask and dissolved in 25 mL H2O. 3 g of potassium thioacetate (1.3 eq, 26.3 mmol) was added to the previous solution and the mixture was heated to 60 °C and stirred overnight. The resulting mixture was concentrated under reduced pressure. The product was extracted by stirring the resulting solid in 100 mL of a solution of [1:1] MeOH/CH2Cl2 at room temperature for 15 min. This process was repeated twice in order to remove KBr. The mixture was filtered, and the filtrate was concentrated under reduced pressure. The expected product 2-(acetothioethyl)-trimethylammonium bromide was formed (4.88 g, 97%) and used without purification. 1H NMR (D2O, 400 MHz): d = 2.27 ppm (3H, s, Ac), 3.04 ppm (9H, s, NMe3), 3.13-3.17 ppm (2H, m, CH2), 3.32-3.37 ppm (2H, m, CH2). 13C NMR (D2O, 100 MHz): d = 21.5, 29.8, 52.8, 64.3 ppm. 1.19 g of the previous product was transferred into a 100 mL round bottom flask and dissolved in 10 mL 6M HCl. The solution was heated to 85 °C and stirred for 1 h. The resulting solution was concentrated under reduced pressure and dried under vacuum. The product 2-mercatoethyl-N,N,N-trimethylammonium chloride was formed (1.12 g, 94%) and used without purification. 1H NMR (D2O, 400 MHz): d = 2.85-2.90 ppm (2H, m, CH2), 3.07 ppm (9H, s, NMe3), 3.12 ppm (1H, s, SH), 3.43-3.49 ppm (2H, m, CH2). 13C NMR (D2O, 100 MHz): d = 16.6, 52.9, 67.8 ppm.

5.7 Amyloid-ß aggregation assay

Lyophilized $A\beta_{42}$ peptide was dissolved in 6 M guanidinium chloride (20 mM NaPi, 200 μM EDTA, pH 8) and incubated for 3 h on ice. This solution was then buffer exchanged into buffer containing 20 mM NaPi and 200 μM EDTA at pH 8 by SEC using a Superdex 75 10/300 GL column (GE Healthcare), and the peak belonging to monomeric $A\beta_{42}$ peptide was isolated and collected in protein low-binding Eppendorf tubes (Corning) on ice. Aggregation reactions were prepared by creating solutions of 1.5 μM $A\beta_{42}$ in the presence of DesAb derivatives in at ratios from 1:2 DesAb to $A\beta_{42}$ and decreasing down to 1:8, ThT was added

to a final concentration of 20 μM. Samples were loaded by pipette into 96-well half-area plates of black polystyrene with clear bottoms and polyethylene glycol coatings (Corning) (80 μL sample/well). Once all samples were loaded plates were sealed to prevent evaporation. Aggregation reactions were carried out at 37 ºC under quiescent conditions using a CLARIOstar plate reader (BMG Labtech). ThT fluorescence was measured every 2 min through the bottom of the plate with an excitation filter of 440 nm and an emissions filter of 480 nm. To create $A\beta_{42}$ fibrils for the seeded assays, 5 μM $A\beta_{42}$ was incubated at 37 ºC under quiescent conditions overnight in protein low-binding Eppendorf tubes (Corning). The fibrils were vigorously pipetted up and down before being added to the aggregation reaction to a final concentration of 30% monomer-equivalents.

5.8 Biolayer interferometry (BLI)

A streptavidin biosensor (ForteBio, Menlo Park, USA) was coated with 5 μg/ml monomeric C-terminally biotinyl ated $A\beta_{42}$ (AnaSpec, Fremont, USA) in PBS + 0.1% BSA (w/v) + 0.02% TWEEN 20 by overnight incubation at 5 °C. Control sensors were coated with the same concentration of biocytin (Sigma). Tips were rinsed before binding experiments by incubation for 1 h at 30 ºC. The association of the antibodies was monitored at concentration of 20 μM in the buffer previously described for 900 s at 30 ºC using an Octet Red96 (ForteBio, Menlo Park, USA). The dissociation was monitored by subsequently placing the sensors into a well containing pure buffer for a further 1200 s. The binding of the antibodies to the biocytin coated streptavidin sensors was subtracted from the signals for immobilized $A\beta_{42}$ to account for non-specific binding. The binding data was analyzed using the Octet software as per the manufacturer's instructions.

5.9 Dha formation on the single-domain antibodies

Cysteine mutants were pre-treated with 100x DTT while shaking at room temperature for 30 minutes, excess DTT was subsequently removed and the protein was buffer exchanged into the reaction buffer (20 mM sodium phosphate buffer (NaPi) at pH 8) by a Zeba Spin desalting column equilibrated with the desired buffer. Protein at concentrations between 75-100 μM was added to 500x solid 2,5-dibromo-1,6-dihexamide (DBHDA, Enamine Ltd.) and left shaking at

37 ºC for 5 hours. The reaction was monitored by LC-MS and when full conversion to Dha was observed excess DBHDA was removed by centrifugation (5 min, 10,000g, 4 ºC).

5.10 Dha formation on Tau

The Tau cysteine mutants were buffer exchanged into 20 mM NaP_i buffer (pH 8) via 7k MWCO Zeba spin desalting columns (Thermo Fisher). 200 µL of 50 µM protein aliquots were reacted with 50 molar equivalents of methyl 2,5-dibromopentanoate (Sigma Aldrich) per cysteine for 12 hours at 37 ºC and shaking at 500 rpm. Excess methyl 2,5-dibromopentanoate was removed by passing the reactions through 7k MWCO Zeba spin desalting columns and then conversion to Dha was verified via LC-MS.

5.11 Circular dichroism

Circular dichroism (CD) spectroscopy was used to analyze protein secondary structure in solution. Antibody samples were diluted to 5 µM in PBS and CD measurements were recorded using a Chirascan spectrophotometer equipped with a Quantum TC125 temperature control unit (25 ºC). The data was acquired in a 0.1 cm path length with a response time of 1 s, a per-point acquisition delay of 5 ms and a pre- and post-scan delay of 50 ms. Spectra were averaged over three scans, in a wavelength range from 200 nm to 250 nm, and the spectrum from a blank sample containing only buffer was subtracted from the averaged data. The structural stability of the proteins was analyzed by monitoring the CD signal at 207 nm from 20 to 98 °C at a rate of 0.5°C min^{-1}. Data points were acquired every 0.1 °C with a bandwidth of 1 nm. Analysis of the thermal unfolding curves was performed, assuming a two-state unfolding model. Tau and its chemical mutants were diluted to 2 µM in 20 mM sodium phosphate buffer and scanned in the same manner except that the range was extended to 195 nm to 250 nm.

5.12 Kinetic analysis

All kinetic analysis was performed using a MatLab script provided from Ref [9]. The time evolution of the total fibril mass concentration, M(t), in the absence of seeds is defined by the integrated rate law

$$\frac{M(t)}{M(\infty)} = 1 - \left(\frac{B_+ + C_+}{B_+ + C_+ e^{kt}} \frac{B_- + C_+ e^{kt}}{B_- + C_+}\right)^{\frac{k_\infty^2}{k\tilde{k}_\infty}} e^{-k_\infty t}$$

where the parameters are described in detail in previous studies[9,85]. These are functions containing the combinations of the two microscopic rate constants k_+k_n and k_+k_2 where k_n, k_+, and k_2 are the primary nucleation, elongation, and secondary nucleation rates respectively. The antibodies can perturb the aggregation reaction by inhibiting one or multiple steps. By applying the above equation to describe the macroscopic profiles we extract the k_+k_n and k_+k_2 parameters that best fit the data to enable accurate comparisons of inhibition mechanisms.

5.13 Tubulin polymerisation assay

The microtubule polymerisations were all performed with reagents from the kit purchased from Cytoskeleton, Inc. (Cat #BK006P) as per manufacturer's instructions. Tubulin as added to the reaction mixture at a concentration of 3 mg/ mL and Tau and the various chemical mutants were assayed at 15 μM for their ability to polymerize the reaction, which was monitored by OD at 340 nM using a CLARIOstar Plus plate reader (BMG Labtech).

5.14 Sanger sequencing

All DNA sequences were analysed by the laboratory of John Lester in the DNA sequencing facility at the Department of Biochemistry (University of Cambridge). All sequencing was performed using a T7 primer for the antibody mutants and a T7-rev primer for the Tau mutants. Sequencing data was analysed using A Plasmid Editor software.

5.15 NanoDrop 2000

All DNA and protein concentrations and quality were measured at 260/280 nm respectively using a NanoDrop 2000 (Thermo Scientific).

5.16 SDS-PAGE

Confirmation of all protein sizes was carried out using pre-cast Novex® NuPAGE® SDS-PAGE gels (Thermo Scientific) with the appropriate standard protein ladders and using NuPAGE® MES SDS running buffer (20x, Thermo Scientific). Electrophoresis was performed at constant voltage 200V for 35 minutes. Photos were taken and analysed using GelDocXR software.

5.17 Common solution recipes

PBS: 1x PBS tablet (Gibco) per 500 mL ddmQ H_2O to yield 10 mM Phosphate, 140 mM NaCl, 2.7 mM KCl, pH 7.4.

LB-Media: 1% (w/v) Tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.0. All LB-media was autoclaved at 121 °C for 15 minutes.

TB-Media: 47.6 g of Terrific broth powder (Sigma Aldrich) was combined with 8 mL of glycerol in 1 L of distilled water. All TB-media was autoclaved at 121 °C for 15 minutes.

LB-Agar: 1.5% Agar (w/v) in LB-media, pH 7.0. All LB-Agar was autoclaved at 121 °C for 15 minutes and poured into plate at <45 °C.

Chapter 6: Bibliography

(1) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.*; 1999; Vol. 13409. https://doi.org/10.1016/S0307-4412(99)00114-4.

(2) Dobson, C. M.; Šali, A.; Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angewandte Chemie - International Edition*. 1998. https://doi.org/10.1002/(SICI)1521-3773(19980420)37:7<868::AID-ANIE868>3.0.CO;2-H.

(3) Dobson, C. M. (CUL-ID:1481556) Protein Folding and Misfolding. *Nature* **2003**, *426* (December), 884–890. https://doi.org/10.1038/nature02261.

(4) Šali, A.; Shakhnovich, E.; Karplus, M. Kinetics of Protein Folding: A Lattice Model Study of the Requirements for Folding to the Native State. *J. Mol. Biol.* **1994**. https://doi.org/10.1006/jmbi.1994.1110.

(5) Dinner, A. R.; Šalib, A.; Smitha, L. J.; Dobsona, C. M.; Karplus, M. Understanding Protein Folding via Free-Energy Surfaces from Theory and Experiment. *Trends in Biochemical Sciences*. 2000. https://doi.org/10.1016/S0968-0004(00)01610-8.

(6) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three Key Residues Form a Critical Contact Network in a Protein Folding Transition State. *Nature* **2001**. https://doi.org/10.1038/35054591.

(7) Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In Vivo Aspects of Protein Folding and Quality Control. *Science*. 2016. https://doi.org/10.1126/science.aac4354.

(8) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. Molecular Chaperones in Protein Folding and Proteostasis. *Nature*. 2011. https://doi.org/10.1038/nature10317.

(9) Arosio, P.; Michaels, T. C. T.; Linse, S.; Månsson, C.; Emanuelsson, C.; Presto, J.; Johansson, J.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Kinetic Analysis Reveals the Diversity of Microscopic Mechanisms through Which Molecular Chaperones Suppress Amyloid Formation. *Nat. Commun.* **2016**, *7*, 10948. https://doi.org/10.1038/ncomms10948.

(10) Fitzpatrick, A. W. P.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Müller, S. A.; et al. Atomic Structure and Hierarchical Assembly of a Cross-β Amyloid Fibril. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (14), 5468–5473. https://doi.org/10.1073/pnas.1219476110.

(11) Fändrich, M.; Fletcher, M. A.; Dobson, C. M. Amyloid Fibrils from Muscle Myoglobin. *Nature* **2001**. https://doi.org/10.1038/35065514.

(12) Eliezer, D.; Yao, J.; Jane Dyson, H.; Wright, P. E. Structural and Dynamic Characterization of Partially Folded States of Apomyoglobin and Implications for Protein Folding. *Nature Structural Biology*. 1998. https://doi.org/10.1038/nsb0298-148.

(13) Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M. Designing Conditions for in Vitro Formation of Amyloid Protofilaments and Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **1999**. https://doi.org/10.1073/pnas.96.7.3590.

(14) Bouchard, M.; Zurdo, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V. Formation of Insulin Amyloid Fibrils Followed by FTIR Simultaneously with CD and Electron Microscopy. *Protein Sci.* **2000**. https://doi.org/10.1110/ps.9.10.1960.

(15) Stroobants, K.; Kumita, J. R.; Harris, N. J.; Chirgadze, D. Y.; Dobson, C. M.; Booth, P. J.; Vendruscolo, M. Amyloid-like Fibrils from an α-Helical Transmembrane Protein. *Biochemistry* **2017**. https://doi.org/10.1021/acs.biochem.7b00157.

(16) Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.;

Frenkel, D.; Gazit, E. Phenylalanine Assembly into Toxic Fibrils Suggests Amyloid Etiology in Phenylketonuria. *Nat. Chem. Biol.* **2012**. https://doi.org/10.1038/nchembio.1002.

(17) Choi, T. B.; Pardridge, W. M. Phenylalanine Transport at the Human Blood-Brain Barrier. Studies with Isolated Human Brain Capillaries. *J. Biol. Chem.* **1986**.

(18) Knowles, T. P.; Fitzpatrick, A. W.; Meehan, S.; Mott, H. R.; Vendruscolo, M.; Dobson, C. M.; Welland, M. E. Role of Intermolecular Forces in Defining Material Properties of Protein Nanofibrils. *Science (80-.).* **2007**, *318* (5858), 1900–1903. https://doi.org/10.1126/science.1150057.

(19) Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W. Functional Amyloid - from Bacteria to Humans. *Trends in Biochemical Sciences*. 2007. https://doi.org/10.1016/j.tibs.2007.03.003.

(20) Fowler, D. M.; Koulov, A. V.; Alory-Jost, C.; Marks, M. S.; Balch, W. E.; Kelly, J. W. Functional Amyloid Formation within Mammalian Tissue. *PLoS Biol.* **2006**. https://doi.org/10.1371/journal.pbio.0040006.

(21) Berson, J. F.; Theos, A. C.; Harper, D. C.; Tenza, D.; Raposo, G.; Marks, M. S. Proprotein Convertase Cleavage Liberates a Fibrillogenic Fragment of a Resident Glycoprotein to Initiate Melanosome Biogenesis. *J. Cell Biol.* **2003**. https://doi.org/10.1083/jcb.200302072.

(22) Berson, J. F.; Harper, D. C.; Tenza, D.; Raposo, G.; Marks, M. S. Pmel17 Initiates Premelanosome Morphogenesis within Multivesicular Bodies. *Mol. Biol. Cell* **2001**. https://doi.org/10.1091/mbc.12.11.3451.

(23) Amaral, M. D. CFTR and Chaperones: Processing and Degradation. *Journal of Molecular Neuroscience*. 2004. https://doi.org/10.1385/jmn:23:1-2:041.

(24) Lomas, D. A.; Carrell, R. W. Serpinopathies and the Conformational Dementias. *Nature Reviews Genetics*. 2002. https://doi.org/10.1038/nrg907.

(25) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75* (1), 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

(26) Sulzer, D. Multiple Hit Hypotheses for Dopamine Neuron Loss in Parkinson's Disease. *Trends in Neurosciences*. 2007. https://doi.org/10.1016/j.tins.2007.03.009.

(27) Fuchs, J.; Nilsson, C.; Kachergus, J.; Munz, M.; Larsson, E. M.; Schüle, B.; Langston, J. W.; Middleton, F. A.; Ross, O. A.; Hulihan, M.; et al. Phenotypic Variation in a Large Swedish Pedigree Due to SNCA Duplication and Triplication. *Neurology* **2007**. https://doi.org/10.1212/01.wnl.0000254458.17630.c5.

(28) Chartier-Harlin, M. C.; Kachergus, J.; Roumier, C.; Mouroux, V.; Douay, X.; Lincoln, S.; Levecque, C.; Larvor, L.; Andrieux, J.; Hulihan, M.; et al. Alpha-Synuclein Locus Duplication as a Cause of Familial Parkinson's Disease. *Lancet* **2004**, *364* (9440), 1167–1169. https://doi.org/10.1016/S0140-6736(04)17103-1.

(29) Sekijima, Y. Transthyretin (ATTR) Amyloidosis: Clinical Spectrum, Molecular Pathogenesis and Disease-Modifying Treatments. *Journal of Neurology, Neurosurgery and Psychiatry*. 2015. https://doi.org/10.1136/jnnp-2014-308724.

(30) Hammarström, P.; Wiseman, R. L.; Powers, E. T.; Kelly, J. W. Prevention of Transthyretin Arnyloid Disease by Changing Protein Misfolding Energetics. *Science (80-.).* **2003**. https://doi.org/10.1126/science.1079589.

(31) Hammarström, P.; Schneider, F.; Kelly, J. W. Trans-Suppression of Misfolding in an Amyloid Disease. *Science (80-.).* **2001**. https://doi.org/10.1126/science.1062245.

(32) Foss, T. R.; Wiseman, R. L.; Kelly, J. W. The Pathway by Which the Tetrameric Protein Transthyretin Dissociates. *Biochemistry* **2005**. https://doi.org/10.1021/bi051608t.

(33) Sekijima, Y.; Kelly, J.; Ikeda, S. Pathogenesis of and Therapeutic Strategies to

Ameliorate the Transthyretin Amyloidoses. *Curr. Pharm. Des.* **2008**. https://doi.org/10.2174/138161208786404155.

(34) Maurer, M. S.; Schwartz, J. H.; Gundapaneni, B.; Elliott, P. M.; Merlini, G.; Waddington-Cruz, M.; Kristen, A. V.; Grogan, M.; Witteles, R.; Damy, T.; et al. Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *N. Engl. J. Med.* **2018**. https://doi.org/10.1056/NEJMoa1805689.

(35) Swaminathan, R.; Ravi, V. K.; Kumar, S.; Kumar, M. V. S.; Chandra, N. Lysozyme: A Model Protein for Amyloid Research. In *Advances in Protein Chemistry and Structural Biology*; 2011. https://doi.org/10.1016/B978-0-12-386483-3.00003-3.

(36) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. Inherent Toxicity of Aggregates Implies a Common Mechanism for Protein Misfolding Diseases. *Nature* **2002**. https://doi.org/10.1038/416507a.

(37) Caughey, B.; Lansbury, P. T. Protofibrils, Pores, Fibrils, and Neurodegeneration: Separating the Responsible Protein Aggregates from the Innocent Bystanders. *Annual Review of Neuroscience*. 2003. https://doi.org/10.1146/annurev.neuro.26.010302.081142.

(38) Winner, B.; Jappelli, R.; Maji, S. K.; Desplats, P. A.; Boyer, L.; Aigner, S.; Hetzer, C.; Loher, T.; Vilar, M.; Campioni, S.; et al. In Vivo Demonstration That α-Synuclein Oligomers Are Toxic. *Proc. Natl. Acad. Sci. U. S. A.* **2011**. https://doi.org/10.1073/pnas.1100976108.

(39) Ross, C. A.; Poirier, M. A. Protein Aggregation and Neurodegenerative Disease. *Nat. Med.* **2004**. https://doi.org/10.1038/nm1066.

(40) Mendes Sousa, M. M.; Cardoso, I.; Fernandes, R.; Guimarães, A.; João Saraiva, M. Deposition of Transthyretin in Early Stages of Familial Amyloidotic Polyneuropathy: Evidence for Toxicity of Nonfibrillar Aggregates. *Am. J. Pathol.* **2001**. https://doi.org/10.1016/S0002-9440(10)63050-7.

(41) Reixach, N.; Deechongkit, S.; Jiang, X.; Kelly, J. W.; Buxbaum, J. N. Tissue Damage in the Amyloidoses: Transthyretin Monomers and Nonnative Oligomers Are the Major Cytotoxic Species in Tissue Culture. *Proc. Natl. Acad. Sci. U. S. A.* **2004**. https://doi.org/10.1073/pnas.0400062101.

(42) Miti, T.; Mulaj, M.; Schmit, J. D.; Muschol, M. Stable, Metastable, and Kinetically Trapped Amyloid Aggregate Phases. *Biomacromolecules* **2015**. https://doi.org/10.1021/bm501521r.

(43) Campioni, S.; Mannini, B.; Zampagni, M.; Pensalfini, A.; Parrini, C.; Evangelisti, E.; Relini, A.; Stefani, M.; Dobson, C. M.; Cecchi, C.; et al. A Causative Link between the Structure of Aberrant Protein Oligomers and Their Toxicity. *Nat. Chem. Biol.* **2010**, *6* (2), 140–147. https://doi.org/10.1038/nchembio.283.

(44) Mannini, B.; Mulvihill, E.; Sgromo, C.; Cascella, R.; Khodarahmi, R.; Ramazzotti, M.; Dobson, C. M.; Cecchi, C.; Chiti, F. Toxicity of Protein Oligomers Is Rationalized by a Function Combining Size and Surface Hydrophobicity. *ACS Chem. Biol.* **2014**. https://doi.org/10.1021/cb500505m.

(45) Fusco, G.; Chen, S. W.; Williamson, P. T. F.; Cascella, R.; Perni, M.; Jarvis, J. A.; Cecchi, C.; Vendruscolo, M.; Chiti, F.; Cremades, N.; et al. Structural Basis of Membrane Disruption and Cellular Toxicity by Alpha-Synuclein Oligomers. *Science (80-.).* **2017**. https://doi.org/10.1126/science.aan6160.

(46) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86* (1), 27–68. https://doi.org/10.1146/annurev-biochem-061516-045115.

(47) Selkoe, D. J.; Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* **2016**. https://doi.org/10.15252/emmm.201606210.

(48) Bateman, R. J.; Xiong, C.; Benzinger, T. L. S.; Fagan, A. M.; Goate, A.; Fox, N. C.; Marcus, D. S.; Cairns, N. J.; Xie, X.; Blazey, T. M.; et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. *N. Engl. J. Med.* **2012**. https://doi.org/10.1056/NEJMoa1202753.

(49) Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Tissue-Based Map of the Human Proteome. *Science (80-.).* **2015**. https://doi.org/10.1126/science.1260419.

(50) Chow, V. W.; Mattson, M. P.; Wong, P. C.; Gleichmann, M. An Overview of APP Processing Enzymes and Products. *Neuromolecular medicine*. 2010. https://doi.org/10.1007/s12017-009-8104-z.

(51) Kelleher, R. J.; Shen, J. Presenilin-1 Mutations and Alzheimer's Disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2017. https://doi.org/10.1073/pnas.1619574114.

(52) Takami, M.; Nagashima, Y.; Sano, Y.; Ishihara, S.; Morishima-Kawashima, M.; Funamoto, S.; Ihara, Y. γ-Secretase: Successive Tripeptide and Tetrapeptide Release from the Transmembrane Domain of β-Carboxyl Terminal Fragment. *J. Neurosci.* **2009**. https://doi.org/10.1523/JNEUROSCI.2362-09.2009.

(53) Li, N.; Liu, K.; Qiu, Y.; Ren, Z.; Dai, R.; Deng, Y.; Qing, H. Effect of Presenilin Mutations on APP Cleavage; Insights into the Pathogenesis of FAD. *Front. Aging Neurosci.* **2016**. https://doi.org/10.3389/fnagi.2016.00051.

(54) Haass, C.; Lemere, C. A.; Capell, A.; Citron, M.; Seubert, P.; Schenk, D.; Lannfelt, L.; Selkoe, D. J. The Swedish Mutation Causes Early-Onset Alzheimer's Disease by β-Secretase Cleavage within the Secretory Pathway. *Nat. Med.* **1995**. https://doi.org/10.1038/nm1295-1291.

(55) Haass, C.; Selkoe, D. J. Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid β-Peptide. *Nature Reviews Molecular Cell Biology*. 2007, pp 101–112. https://doi.org/10.1038/nrm2101.

(56) Meli, G.; Visintin, M.; Cannistraci, I.; Cattaneo, A. Direct in Vivo Intracellular Selection of Conformation-Sensitive Antibody Domains Targeting Alzheimer's Amyloid-β Oligomers. *J. Mol. Biol.* **2009**. https://doi.org/10.1016/j.jmb.2009.01.061.

(57) Scopa, C.; Marrocco, F.; Latina, V.; Ruggeri, F.; Corvaglia, V.; La Regina, F.; Ammassari-Teule, M.; Middei, S.; Amadoro, G.; Meli, G.; et al. Impaired Adult Neurogenesis Is an Early Event in Alzheimer's Disease Neurodegeneration, Mediated by Intracellular Aβ Oligomers. *Cell Death Differ.* **2020**. https://doi.org/10.1038/s41418-019-0409-3.

(58) Heneka, M. T.; Carson, M. J.; Khoury, J. El; Landreth, G. E.; Brosseron, F.; Feinstein, D. L.; Jacobs, A. H.; Wyss-Coray, T.; Vitorica, J.; Ransohoff, R. M.; et al. Neuroinflammation in Alzheimer's Disease. *The Lancet Neurology*. 2015. https://doi.org/10.1016/S1474-4422(15)70016-5.

(59) Stephan, A. H.; Barres, B. A.; Stevens, B. The Complement System: An Unexpected Role in Synaptic Pruning during Development and Disease. *Annual Review of Neuroscience*. 2012. https://doi.org/10.1146/annurev-neuro-061010-113810.

(60) Hong, S.; Beja-Glasser, V. F.; Nfonoyim, B. M.; Frouin, A.; Li, S.; Ramakrishnan, S.; Merry, K. M.; Shi, Q.; Rosenthal, A.; Barres, B. A.; et al. Complement and Microglia Mediate Early Synapse Loss in Alzheimer Mouse Models. *Science (80-.).* **2016**. https://doi.org/10.1126/science.aad8373.

(61) Mucke, L.; Masliah, E.; Yu, G. Q.; Mallory, M.; Rockenstein, E. M.; Tatsuno, G.; Hu, K.; Kholodenko, D.; Johnson-Wood, K.; McConlogue, L. High-Level Neuronal Expression of Aβ(1-42) in Wild-Type Human Amyloid Protein Precursor Transgenic Mice: Synaptotoxicity without Plaque Formation. *J. Neurosci.* **2000**.

https://doi.org/10.1523/jneurosci.20-11-04050.2000.
(62) Selkoe, D. J. Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiological Reviews*. 2001. https://doi.org/10.1152/physrev.2001.81.2.741.
(63) Liddelow, S. A.; Guttenplan, K. A.; Clarke, L. E.; Bennett, F. C.; Bohlen, C. J.; Schirmer, L.; Bennett, M. L.; Münch, A. E.; Chung, W. S.; Peterson, T. C.; et al. Neurotoxic Reactive Astrocytes Are Induced by Activated Microglia. *Nature* **2017**. https://doi.org/10.1038/nature21029.
(64) Wang, Y.; Mandelkow, E. Tau in Physiology and Pathology. *Nature Reviews Neuroscience*. 2016. https://doi.org/10.1038/nrn.2015.1.
(65) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. *PLoS Biol.* **2009**. https://doi.org/10.1371/journal.pbio.1000034.
(66) Götz, J.; Halliday, G.; Nisbet, R. M. Molecular Pathogenesis of the Tauopathies. *Annual Review of Pathology: Mechanisms of Disease*. 2019. https://doi.org/10.1146/annurev-pathmechdis-012418-012936.
(67) Falcon, B.; Zhang, W.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Ghetti, B.; Scheres, S. H. W.; Goedert, M. Structures of Filaments from Pick's Disease Reveal a Novel Tau Protein Fold. *Nature* **2018**. https://doi.org/10.1038/s41586-018-0454-y.
(68) Von Bergen, M.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E. M.; Mandelkow, E. Assembly of τ Protein into Alzheimer Paired Helical Filaments Depends on a Local Sequence Motif (306VQIVYK311) Forming β Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2000**. https://doi.org/10.1073/pnas.97.10.5129.
(69) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; et al. Atomic Structures of Amyloid Cross-β Spines Reveal Varied Steric Zippers. *Nature* **2007**. https://doi.org/10.1038/nature05695.
(70) Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. *Nature* **2017**. https://doi.org/10.1038/nature23002.
(71) Hanger, D. P.; Anderton, B. H.; Noble, W. Tau Phosphorylation: The Therapeutic Challenge for Neurodegenerative Disease. *Trends in Molecular Medicine*. 2009. https://doi.org/10.1016/j.molmed.2009.01.003.
(72) Kadavath, H.; Hofele, R. V.; Biernat, J.; Kumar, S.; Tepper, K.; Urlaub, H.; Mandelkow, E.; Zweckstetter, M. Tau Stabilizes Microtubules by Binding at the Interface between Tubulin Heterodimers. *Proc. Natl. Acad. Sci. U. S. A.* **2015**. https://doi.org/10.1073/pnas.1504081112.
(73) Gianni, S.; Camilloni, C.; Giri, R.; Toto, A.; Bonetti, D.; Morrone, A.; Sormanni, P.; Brunori, M.; Vendruscolo, M. Understanding the Frustration Arising from the Competition between Function, Misfolding, and Aggregation in a Globular Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014**. https://doi.org/10.1073/pnas.1405233111.
(74) Lewis, J.; Dickson, D. W.; Lin, W. L.; Chisholm, L.; Corral, A.; Jones, G.; Yen, S. H.; Sahara, N.; Skipper, L.; Yager, D.; et al. Enhanced Neurofibrillary Degeneration in Transgenic Mice Expressing Mutant Tau and APP. *Science (80-.).* **2001**. https://doi.org/10.1126/science.1058189.
(75) Jin, M.; Shepardson, N.; Yang, T.; Chen, G.; Walsh, D.; Selkoe, D. J. Soluble Amyloid β-Protein Dimers Isolated from Alzheimer Cortex Directly Induce Tau Hyperphosphorylation and Neuritic Degeneration. *Proc. Natl. Acad. Sci. U. S. A.* **2011**.

https://doi.org/10.1073/pnas.1017033108.
(76) Palavicini, J. P.; Wang, C.; Chen, L.; Hosang, K.; Wang, J.; Tomiyama, T.; Mori, H.; Han, X. Oligomeric Amyloid-Beta Induces MAPK-Mediated Activation of Brain Cytosolic and Calcium-Independent Phospholipase A2 in a Spatial-Specific Manner. *Acta Neuropathol. Commun.* **2017**. https://doi.org/10.1186/s40478-017-0460-6.
(77) Lee, I. S.; Jung, K.; Kim, I. S.; Park, K. I. Amyloid-β Oligomers Regulate the Properties of Human Neural Stem Cells through GSK-3β Signaling. *Exp. Mol. Med.* **2013**. https://doi.org/10.1038/emm.2013.125.
(78) Zheng, W. H.; Bastianetto, S.; Mennicken, F.; Ma, W.; Kar, S. Amyloid β Peptide Induces Tau Phosphorylation and Loss of Cholinergic Neurons in Rat Primary Septal Cultures. *Neuroscience* **2002**. https://doi.org/10.1016/S0306-4522(02)00404-9.
(79) Thies, E.; Mandelkow, E. M. Missorting of Tau in Neurons Causes Degeneration of Synapses That Can Be Rescued by the Kinase MARK2/Par-1. *J. Neurosci.* **2007**. https://doi.org/10.1523/JNEUROSCI.4674-06.2007.
(80) Guillozet-Bongaarts, A. L.; Cahill, M. E.; Cryns, V. L.; Reynolds, M. R.; Berry, R. W.; Binder, L. I. Pseudophosphorylation of Tau at Serine 422 Inhibits Caspase Cleavage: In Vitro Evidence and Implications for Tangle Formation in Vivo. *J. Neurochem.* **2006**. https://doi.org/10.1111/j.1471-4159.2006.03784.x.
(81) Dickey, C. A.; Kamal, A.; Lundgren, K.; Klosak, N.; Bailey, R. M.; Dunmore, J.; Ash, P.; Shoraka, S.; Zlatkovic, J.; Eckman, C. B.; et al. The High-Affinity HSP90-CHIP Complex Recognizes and Selectively Degrades Phosphorylated Tau Client Proteins. *J. Clin. Invest.* **2007**. https://doi.org/10.1172/JCI29715.
(82) Dixit, R.; Ross, J. L.; Goldman, Y. E.; Holzbaur, E. L. F. Differential Regulation of Dynein and Kinesin Motor Proteins by Tau. *Science (80-.).* **2008**. https://doi.org/10.1126/science.1152993.
(83) Morris, M.; Knudsen, G. M.; Maeda, S.; Trinidad, J. C.; Ioanoviciu, A.; Burlingame, A. L.; Mucke, L. Tau Post-Translational Modifications in Wild-Type and Human Amyloid Precursor Protein Transgenic Mice. *Nat. Neurosci.* **2015**. https://doi.org/10.1038/nn.4067.
(84) Habchi, J.; Chia, S.; Galvagnion, C.; Michaels, T. C. T.; Bellaiche, M. M. J.; Ruggeri, F. S.; Sanguanini, M.; Idini, I.; Kumita, J. R.; Sparr, E.; et al. Cholesterol Catalyses Aβ42 Aggregation through a Heterogeneous Nucleation Pathway in the Presence of Lipid Membranes. *Nat. Chem.* **2018**. https://doi.org/10.1038/s41557-018-0031-x.
(85) Michaels, T. C. T.; Šarić, A.; Habchi, J.; Chia, S.; Meisl, G.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Chemical Kinetics for Bridging Molecular Mechanisms and Macroscopic Measurements of Amyloid Fibril Formation. *Annu. Rev. Phys. Chem.* **2018**. https://doi.org/10.1146/annurev-physchem-050317-021322.
(86) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science (80-.).* **2009**. https://doi.org/10.1126/science.1178250.
(87) Koffie, R. M.; Meyer-Luehmann, M.; Hashimoto, T.; Adams, K. W.; Mielke, M. L.; Garcia-Alloza, M.; Micheva, K. D.; Smith, S. J.; Kim, M. L.; Lee, V. M.; et al. Oligomeric Amyloid β Associates with Postsynaptic Densities and Correlates with Excitatory Synapse Loss near Senile Plaques. *Proc. Natl. Acad. Sci. U. S. A.* **2009**. https://doi.org/10.1073/pnas.0811698106.
(88) Arosio, P.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Chemical Kinetics for Drug Discovery to Combat Protein Aggregation Diseases. *Trends Pharmacol. Sci.* **2014**, *35* (3), 127–135. https://doi.org/10.1016/j.tips.2013.12.005.
(89) Chia, S.; Habchi, J.; Michaels, T. C. T.; Cohen, S. I. A.; Linse, S.; Dobson, C. M.;

Knowles, T. P. J.; Vendruscolo, M. SAR by Kinetics for Drug Discovery in Protein Misfolding Diseases. *Proc. Natl. Acad. Sci. U. S. A.* **2018**. https://doi.org/10.1073/pnas.1807884115.

(90) Perni, M.; Galvagnion, C.; Maltsev, A.; Meisl, G.; Müller, M. B. D.; Challa, P. K.; Kirkegaard, J. B.; Flagmeier, P.; Cohen, S. I. A.; Cascella, R.; et al. A Natural Product Inhibits the Initiation of α-Synuclein Aggregation and Suppresses Its Toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 201610586. https://doi.org/10.1073/pnas.1610586114.

(91) Scott, A. M.; Wolchok, J. D.; Old, L. J. Antibody Therapy of Cancer. *Nature Reviews Cancer*. 2012. https://doi.org/10.1038/nrc3236.

(92) Trail, P. Antibody Drug Conjugates as Cancer Therapeutics. *Antibodies* **2013**. https://doi.org/10.3390/antib2010113.

(93) Chozinski, T. J.; Halpern, A. R.; Okawa, H.; Kim, H. J.; Tremel, G. J.; Wong, R. O. L.; Vaughan, J. C. Expansion Microscopy with Conventional Antibodies and Fluorescent Proteins. *Nat. Methods* **2016**. https://doi.org/10.1038/nmeth.3833.

(94) Carmon, K. S.; Azhdarinia, A. Application of Immuno-PET in Antibody–Drug Conjugate Development. *Molecular Imaging*. 2018. https://doi.org/10.1177/1536012118801223.

(95) Hilvert, D. Critical Analysis of Antibody Catalysis. *Annual Review of Biochemistry*. 2000. https://doi.org/10.1146/annurev.biochem.69.1.751.

(96) Harris, L. J.; Larson, S. B.; Hasel, K. W.; McPherson, A. Refined Structure of an Intact IgG2a Monoclonal Antibody. *Biochemistry* **1997**. https://doi.org/10.1021/bi962514+.

(97) Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **2013**. https://doi.org/10.1146/annurev-biochem-063011-092449.

(98) Flajnik, M. F.; Kasahara, M. Origin and Evolution of the Adaptive Immune System: Genetic Events and Selective Pressures. *Nature Reviews Genetics*. 2010. https://doi.org/10.1038/nrg2703.

(99) Barthelemy, P. A.; Raab, H.; Appleton, B. A.; Bond, C. J.; Wu, P.; Wiesmann, C.; Sidhu, S. S. Comprehensive Analysis of the Factors Contributing to the Stability and Solubility of Autonomous Human VH Domains. *J. Biol. Chem.* **2008**. https://doi.org/10.1074/jbc.M708536200.

(100) Desmyter, A.; Decanniere, K.; Muyldermans, S.; Wyns, L. Antigen Specificity and High Affinity Binding Provided by One Single Loop of a Camel Single-Domain Antibody. *J. Biol. Chem.* **2001**. https://doi.org/10.1074/jbc.M102107200.

(101) Desmyter, A.; Transue, T. R.; Ghahroudi, M. A.; Thi, M. H. D.; Poortmans, F.; Hamers, R.; Muyldermans, S.; Wyns, L. Crystal Structure of a Camel Single-Domain V(H) Antibody Fragment in Complex with Lysozyme. *Nat. Struct. Biol.* **1996**. https://doi.org/10.1038/nsb0996-803.

(102) Li, T.; Bourgeois, J. P.; Celli, S.; Glacial, F.; Le Sourd, A. M.; Mecheri, S.; Weksler, B.; Romero, I.; Couraud, P. O.; Rougeon, F.; et al. Cell-Penetrating Anti-GFAP VHH and Corresponding Fluorescent Fusion Protein VHH-GFP Spontaneously Cross the Blood-Brain Barrier and Specifically Recognize Astrocytes: Application to Brain Imaging. *FASEB J.* **2012**. https://doi.org/10.1096/fj.11-201384.

(103) Roopenian, D. C.; Akilesh, S. FcRn: The Neonatal Fc Receptor Comes of Age. *Nature Reviews Immunology*. 2007. https://doi.org/10.1038/nri2155.

(104) Wagner, S. D.; Neuberger, M. S. Somatic Hypermutation of Immunoglobulin Genes. *Annual Review of Immunology*. 1996. https://doi.org/10.1146/annurev.immunol.14.1.441.

(105) Childs, L. M.; Baskerville, E. B.; Cobey, S. Trade-Offs in Antibody Repertoires to Complex Antigens. *Philos. Trans. R. Soc. B Biol. Sci.* **2015**. https://doi.org/10.1098/rstb.2014.0245.

(106) Merle, N. S.; Noe, R.; Halbwachs-Mecarelli, L.; Fremeaux-Bacchi, V.; Roumenina, L. T. Complement System Part II: Role in Immunity. *Frontiers in Immunology*. 2015. https://doi.org/10.3389/fimmu.2015.00257.
(107) Li, Y.; O'Dell, S.; Walker, L. M.; Wu, X.; Guenaga, J.; Feng, Y.; Schmidt, S. D.; McKee, K.; Louder, M. K.; Ledgerwood, J. E.; et al. Mechanism of Neutralization by the Broadly Neutralizing HIV-1 Monoclonal Antibody VRC01. *J. Virol.* **2011**. https://doi.org/10.1128/jvi.00754-11.
(108) Fabrizio, K.; Manix, C.; Guimaraeš, A. J.; Nosanchuk, J. D.; Pirofski, L. A. Aggregation of Streptococcus Pneumoniae by a Pneumococcal Capsular Polysaccharide-Specific Human Monoclonal IgM Correlates with Antibody Efficacy In Vivo. *Clin. Vaccine Immunol.* **2010**. https://doi.org/10.1128/CVI.00410-09.
(109) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nature Reviews Genetics*. 2015. https://doi.org/10.1038/nrg3927.
(110) Cherf, G. M.; Cochran, J. R. Applications of Yeast Surface Display for Protein Engineering. *Methods Mol. Biol.* **2015**. https://doi.org/10.1007/978-1-4939-2748-7_8.
(111) Takahashi, T. T.; Austin, R. J.; Roberts, R. W. MRNA Display: Ligand Discovery, Interaction Analysis and Beyond. *Trends in Biochemical Sciences*. 2003. https://doi.org/10.1016/S0968-0004(03)00036-7.
(112) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. Third Generation Antibody Discovery Methods:: In Silico Rational Design. *Chemical Society Reviews*. 2018. https://doi.org/10.1039/c8cs00523k.
(113) Zheng, T.; Xie, J.; Yang, Z.; Tao, P.; Shi, B.; Douthit, L.; Wu, P.; Lerner, R. A. Antibody Selection Using Clonal Cocultivation of Escherichia Coli and Eukaryotic Cells in Miniecosystems. *Proc. Natl. Acad. Sci. U. S. A.* **2018**. https://doi.org/10.1073/pnas.1806718115.
(114) Guy, A. J.; Irani, V.; MacRaild, C. A.; Anders, R. F.; Norton, R. S.; Beeson, J. G.; Richards, J. S.; Ramsland, P. A. Insights into the Immunological Properties of Intrinsically Disordered Malaria Proteins Using Proteome Scale Predictions. *PLoS One* **2015**. https://doi.org/10.1371/journal.pone.0141729.
(115) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. Rational Design of Antibodies Targeting Specific Epitopes within Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (32), 9902–9907. https://doi.org/10.1073/pnas.1422401112.
(116) Julian, M. C.; Li, L.; Garde, S.; Wilen, R.; Tessier, P. M. Efficient Affinity Maturation of Antibody Variable Domains Requires Co-Selection of Compensatory Mutations to Maintain Thermodynamic Stability. *Sci. Rep.* **2017**. https://doi.org/10.1038/srep45259.
(117) Perchiacca, J. M.; Ladiwala, A. R. A.; Bhattacharya, M.; Tessier, P. M. Structure-Based Design of Conformation- and Sequence-Specific Antibodies against Amyloid β. *Proc. Natl. Acad. Sci. U. S. A.* **2012**. https://doi.org/10.1073/pnas.1111232108.
(118) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **2015**, *427* (2), 478–490. https://doi.org/10.1016/j.jmb.2014.09.026.
(119) Jarasch, A.; Koll, H.; Regula, J. T.; Bader, M.; Papadimitriou, A.; Kettenberger, H. Developability Assessment during the Selection of Novel Therapeutic Antibodies. *J. Pharm. Sci.* **2015**. https://doi.org/10.1002/jps.24430.
(120) Kayed, R.; Canto, I.; Breydo, L.; Rasool, S.; Lukacsovich, T.; Wu, J.; Albay, R.; Pensalfini, A.; Yeung, S.; Head, E.; et al. Conformation Dependent Monoclonal Antibodies Distinguish Different Replicating Strains or Conformers of Prefibrillar Aβ Oligomers. *Mol. Neurodegener.* **2010**. https://doi.org/10.1186/1750-1326-5-57.
(121) van Dyck, C. H. Anti-Amyloid-β Monoclonal Antibodies for Alzheimer's Disease: Pitfalls and Promise. *Biological Psychiatry*. 2018.

https://doi.org/10.1016/j.biopsych.2017.08.010.
(122) Gilman, S.; Koller, M.; Black, R. S.; Jenkins, L.; Griffith, S. G.; Fox, N. C.; Eisner, L.; Kirby, L.; Boada Rovira, M.; Forette, F.; et al. Clinical Effects of Aβ Immunization (AN1792) in Patients with AD in an Interrupted Trial. *Neurology* **2005**. https://doi.org/10.1212/01.WNL.0000159740.16984.3C.
(124) Montoliu-Gaya, L.; Villegas, S. Aβ-Immunotherapeutic Strategies: A Wide Range of Approaches for Alzheimer's Disease Treatment. *Expert Rev. Mol. Med.* **2016**. https://doi.org/10.1017/erm.2016.11.
(125) Black, R. S.; Sperling, R. A.; Safirstein, B.; Motter, R. N.; Pallay, A.; Nichols, A.; Grundman, M. A Single Ascending Dose Study of Bapineuzumab in Patients with Alzheimer Disease. *Alzheimer Dis. Assoc. Disord.* **2010**. https://doi.org/10.1097/WAD.0b013e3181c53b00.
(126) Selkoe, D. J. Resolving Controversies on the Path to Alzheimer's Therapeutics. *Nature Medicine*. 2011. https://doi.org/10.1038/nm.2460.
(127) The Lancet Neurology. Solanezumab: Too Late in Mild Alzheimer's Disease? *The Lancet Neurology*. 2017. https://doi.org/10.1016/S1474-4422(16)30395-7.
(128) Schneider, L. A Resurrection of Aducanumab for Alzheimer's Disease. *Lancet Neurol.* **2020**. https://doi.org/10.1016/S1474-4422(19)30480-6.
(129) Sevigny, J.; Chiao, P.; Bussière, T.; Weinreb, P. H.; Williams, L.; Maier, M.; Dunstan, R.; Salloway, S.; Chen, T.; Ling, Y.; et al. The Antibody Aducanumab Reduces Aβ Plaques in Alzheimer's Disease. *Nature* **2016**. https://doi.org/10.1038/nature19323.
(130) Linse, S. S.; Scheidt, T.; Bernfur, K.; Vendruscolo, M.; Dobson, C. M.; Cohen, S. I. A.; Sileikis, E.; Lundquist, M.; Qian, F.; O'Malley, T. T.; et al. Kinetic Fingerprint of Antibody Therapies Predicts Outcomes of Alzheimer Clinical Trials. *bioRxiv* **2019**.https://doi.org/10.1093/hmg/10.14.1511.
(131) Rodriguez, E. A.; Campbell, R. E.; Lin, J. Y.; Lin, M. Z.; Miyawaki, A.; Palmer, A. E.; Shu, X.; Zhang, J.; Tsien, R. Y. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends in Biochemical Sciences*. 2017. https://doi.org/10.1016/j.tibs.2016.09.010.
(132) Prescher, J. A.; Bertozzi, C. R. Chemistry in Living Systems. *Nature Chemical Biology*. 2005. https://doi.org/10.1038/nchembio0605-13.
(133) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie - International Edition*. 2009. https://doi.org/10.1002/anie.200900942.
(134) Kottegoda, S.; Aoto, P. C.; Sims, C. E.; Allbritton, N. L. Biarsenical-Tetracysteine Motif as a Fluorescent Tag for Detection in Capillary Electrophoresis. *Anal. Chem.* **2008**. https://doi.org/10.1021/ac8003242.
(135) Guignet, E. G.; Hovius, R.; Vogel, H. Reversible Site-Selective Labeling of Membrane Proteins in Live Cells. *Nat. Biotechnol.* **2004**, *22* (4), 440–444. https://doi.org/10.1038/nbt954.
(136) Giriat, I.; Muir, T. W. Protein Semi-Synthesis in Living Cells. *J. Am. Chem. Soc.* **2003**. https://doi.org/10.1021/ja034736i.
(137) Akkapeddi, P.; Azizi, S.-A.; Freedy, A. M.; Cal, P. M. S. D.; Gois, P. M. P.; Bernardes, G. J. L. Construction of Homogeneous Antibody–Drug Conjugates Using Site-Selective Protein Chemistry. *Chem. Sci.* **2016**, *7* (5), 2954–2963. https://doi.org/10.1039/C6SC00170J.
(138) Chapman, A. P. PEGylated Antibodies and Antibody Fragments for Improved Therapy: A Review. *Adv. Drug Deliv. Rev.* **2002**. https://doi.org/10.1016/S0169-409X(02)00026-1.
(139) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. A "Tag-and-Modify" Approach to Site-

Selective Protein Modification. *Acc. Chem. Res.* **2011**. https://doi.org/10.1021/ar200056q.
(140) Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L. Contemporary Approaches to Site-Selective Protein Modification. *Nature Reviews Chemistry*. 2019. https://doi.org/10.1038/s41570-019-0079-1.
(141) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews*. 2014, pp 4764–4806. https://doi.org/10.1021/cr400355w.
(142) Cowie, D. B.; Cohen, G. N.; Bolton, E. T.; De Robichon-Szulmajster, H. Amino Acid Analog Incorporation into Bacterial Proteins. *BBA - Biochim. Biophys. Acta* **1959**. https://doi.org/10.1016/0006-3002(59)90230-6.
(143) Tang, Y.; Tirrell, D. A. Attenuation of the Editing Activity of the Escherichia Coli Leucyl-TRNA Synthetase Allows Incorporation of Novel Amino Acids into Proteins in Vivo. *Biochemistry* **2002**. https://doi.org/10.1021/bi026130x.
(144) Wang, P.; Vaidehi, N.; Tirrell, D. A.; Goddard, W. A. Virtual Screening for Binding of Phenylalanine Analogues to Phenylalanyl-TRNA Synthetase. *J. Am. Chem. Soc.* **2002**. https://doi.org/10.1021/ja0175441.
(145) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective Identification of Newly Synthesized Proteins in Mammalian Cells Using Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). *Proc. Natl. Acad. Sci. U. S. A.* **2006**. https://doi.org/10.1073/pnas.0601637103.
(146) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**. https://doi.org/10.1146/annurev.biochem.052308.105824.
(147) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science (80-.).* **2001**. https://doi.org/10.1126/science.1060077.
(148) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. Inverse Electron Demand Diels-Alder Reactions in Chemical Biology. *Chemical Society Reviews*. 2017. https://doi.org/10.1039/c7cs00184c.
(149) Rossin, R.; Versteegen, R. M.; Wu, J.; Khasanov, A.; Wessels, H. J.; Steenbergen, E. J.; Ten Hoeve, W.; Janssen, H. M.; Van Onzen, A. H. A. M.; Hudson, P. J.; et al. Chemically Triggered Drug Release from an Antibody-Drug Conjugate Leads to Potent Antitumour Activity in Mice. *Nat. Commun.* **2018**. https://doi.org/10.1038/s41467-018-03880-y.
(150) Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beránek, V.; et al. Total Synthesis of Escherichia Coli with a Recoded Genome. *Nature* **2019**. https://doi.org/10.1038/s41586-019-1192-5.
(151) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* **2010**. https://doi.org/10.1038/nature08817.
(152) Zioudrou, C.; Wilchek, M.; Patchornik, A. Conversion of the L-Serine Residue to an L-Cysteine Residue in Peptides. *Biochemistry* **1965**. https://doi.org/10.1021/bi00885a018.
(153) Neet, K. E.; Koshland, D. E. The Conversion of Serine at the Active Site of Subtilisin to Cysteine: A "Chemical Mutation". *Proc. Natl. Acad. Sci. U. S. A.* **1966**. https://doi.org/10.1073/pnas.56.5.1606.
(154) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**. https://doi.org/10.1021/ja003265m.
(155) Crich, D.; Banerjee, A. Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* **2007**. https://doi.org/10.1021/ja072804l.

(156) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G. A Genetically Encoded Boronate-Containing Amino Acid. *Angew. Chemie - Int. Ed.* **2008**. https://doi.org/10.1002/anie.200803240.
(157) Li, B.; Yu, J. P. J.; Brunzelle, J. S.; Moll, G. N.; Van Der Donk, W. A.; Nair, S. K. Structure and Mechanism of the Lantibiotic Cyclase Involved in Nisin Biosynthesis. *Science (80-.).* **2006**. https://doi.org/10.1126/science.1121422.
(158) Wang, J.; Schiller, S. M.; Schultz, P. G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chemie - Int. Ed.* **2007**. https://doi.org/10.1002/anie.200702305.
(159) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. Methods for Converting Cysteine to Dehydroalanine on Peptides and Proteins. *Chem. Sci.* **2011**, *2* (9), 1666. https://doi.org/10.1039/c1sc00185j.
(160) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chemie - Int. Ed.* **2012**. https://doi.org/10.1002/anie.201106432.
(161) Freedy, A. M.; Matos, M. J.; Boutureira, O.; Corzana, F.; Guerreiro, A.; Akkapeddi, P.; Somovilla, V. J.; Rodrigues, T.; Nicholls, K.; Xie, B.; et al. Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation. *J. Am. Chem. Soc.* **2017**. https://doi.org/10.1021/jacs.7b10702.
(162) Dadová, J.; Wu, K. J.; Isenegger, P. G.; Errey, J. C.; Bernardes, G. J. L.; Chalker, J. M.; Raich, L.; Rovira, C.; Davis, B. G. Precise Probing of Residue Roles by Post-Translational β,γ-C,N Aza-Michael Mutagenesis in Enzyme Active Sites. *ACS Cent. Sci.* **2017**. https://doi.org/10.1021/acscentsci.7b00341.
(163) Chooi, K. P.; Galan, S. R. G.; Raj, R.; McCullagh, J.; Mohammed, S.; Jones, L. H.; Davis, B. G. Synthetic Phosphorylation of P38α Recapitulates Protein Kinase Activity. *J. Am. Chem. Soc.* **2014**. https://doi.org/10.1021/ja4095318.
(164) Diskin, M.; Askari, N.; Capone, R.; Engelberg, D.; Livnah, O. Active Mutants of the Human P38α Mitogen-Activated Protein Kinase. *J. Biol. Chem.* **2004**. https://doi.org/10.1074/jbc.M404595200.
(165) Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W. L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; et al. Posttranslational Mutagenesis: A Chemical Strategy for Exploring Protein Side-Chain Diversity. *Science (80-.).* **2016**. https://doi.org/10.1126/science.aag1465.
(166) Andricopulo, A.; Montanari, C. Structure-Activity Relationships for the Design of Small-Molecule Inhibitors. *Mini-Reviews Med. Chem.* **2005**. https://doi.org/10.2174/1389557054023224.
(167) Liu, C. C.; Mack, A. V.; Tsao, M. L.; Mills, J. H.; Hyun, S. L.; Choe, H.; Farzan, M.; Schultz, P. G.; Smider, V. V. Protein Evolution with an Expanded Genetic Code. *Proc. Natl. Acad. Sci. U. S. A.* **2008**. https://doi.org/10.1073/pnas.0809543105.
(168) Windle, C. L.; Simmons, K. J.; Ault, J. R.; Trinh, C. H.; Nelson, A.; Pearson, A. R.; Berry, A. Extending Enzyme Molecular Recognition with an Expanded Amino Acid Alphabet. *Proc. Natl. Acad. Sci. U. S. A.* **2017**. https://doi.org/10.1073/pnas.1616816114.
(169) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**. https://doi.org/10.1186/1472-6750-8-91.
(170) Eidenmuller, J.; Fath, T.; Hellwig, A.; Reed, J.; Sontag, E.; Brandt, R. Structural and Functional Implications of Tau Hyperphosphorylation: Information from

Phosphorylation-Mimicking Mutated Tau Proteins. *Biochemistry* **2000**. https://doi.org/10.1021/bi001290z.
(171) Yoshida, H.; Goedert, M. Phosphorylation of Microtubule-Associated Protein Tau by AMPK-Related Kinases. *J. Neurochem.* **2012**. https://doi.org/10.1111/j.1471-4159.2011.07523.x.
(172) Mulot, S. F. C.; Hughes, K.; Woodgett, J. R.; Anderton, B. H.; Hanger, D. P. PHF-Tau from Alzheimer's Brain Comprises Four Species on SDS-PAGE Which Can Be Mimicked by in Vitro Phosphorylation of Human Brain Tau by Glycogen Synthase Kinase-3β. *FEBS Lett.* **1994**. https://doi.org/10.1016/0014-5793(94)00702-0.
(173) Haj-Yahya, M.; Lashuel, H. A. Protein Semisynthesis Provides Access to Tau Disease-Associated Post-Translational Modifications (PTMs) and Paves the Way to Deciphering the Tau PTM Code in Health and Diseased States. *J. Am. Chem. Soc.* **2018**. https://doi.org/10.1021/jacs.8b02668.
(174) Bilyard, M. K.; Bailey, H. J.; Raich, L.; Gafitescu, M. A.; Machida, T.; Iglésias-Fernández, J.; Lee, S. S.; Spicer, C. D.; Rovira, C.; Yue, W. W.; et al. Palladium-Mediated Enzyme Activation Suggests Multiphase Initiation of Glycogenesis. *Nature* **2018**. https://doi.org/10.1038/s41586-018-0644-7.
(175) Haj-Yahya, M.; Gopinath, P.; Rajasekhar, K.; Mirbaha, H.; Diamond, M. I.; Lashuel, H. A. Site-Specific Hyperphosphorylation Inhibits, Rather than Promotes, Tau Fibrillization, Seeding Capacity, and Its Microtubule Binding. *Angew. Chemie - Int. Ed.* **2020**. https://doi.org/10.1002/anie.201913001.
(176) Siddhartha, M. R.; Anahi, S. G.; Perla, G. P.; Macías, M.; Ordaz, B.; Fernando, P. O.; Azucena, A. V.; Erika, O. S.; Sofía, D. C.; Perry, G.; et al. Phosphorylation of Tau Protein Correlates with Changes in Hippocampal Theta Oscillations and Reduces Hippocampal Excitability in Alzheimer's Model. *J. Biol. Chem.* **2018**. https://doi.org/10.1074/jbc.RA117.001187.
(177) Hampel, H.; Buerger, K.; Zinkowski, R.; Teipel, S. J.; Goernitz, A.; Andreasen, N.; Sjoegren, M.; DeBernardis, J.; Kerkman, D.; Ishiguro, K.; et al. Measurement of Phosphorylated Tau Epitopes in the Differential Diagnosis of Alzheimer Disease: A Comparative Cerebrospinal Fluid Study. *Arch. Gen. Psychiatry* **2004**. https://doi.org/10.1001/archpsyc.61.1.95.
(178) Kundel, F.; De, S.; Flagmeier, P.; Horrocks, M. H.; Kjaergaard, M.; Shammas, S. L.; Jackson, S. E.; Dobson, C. M.; Klenerman, D. Hsp70 Inhibits the Nucleation and Elongation of Tau and Sequesters Tau Aggregates with High Affinity. *ACS Chem. Biol.* **2018**. https://doi.org/10.1021/acschembio.7b01039.
(179) Morrison, P. M.; Foley, P. J.; Warriner, S. L.; Webb, M. E. Chemical Generation and Modification of Peptides Containing Multiple Dehydroalanines. *Chem. Commun.* **2015**. https://doi.org/10.1039/c5cc05469a.
(180) Smith, M. J.; Crowther, R. A.; Goedert, M. The Natural Osmolyte Trimethylamine N-Oxide (TMAO) Restores the Ability of Mutant Tau to Promote Microtubule Assembly. *FEBS Lett.* **2000**. https://doi.org/10.1016/S0014-5793(00)02169-4.
(181) Kanai, Y.; Chen, J.; Hirokawa, N. Microtubule Bundling by Tau Proteins in Vivo: Analysis of Functional Domains. *EMBO J.* **1992**. https://doi.org/10.1002/j.1460-2075.1992.tb05489.x.
(182) McKibben, K. M.; Rhoades, E. Independent Tubulin Binding and Polymerization by the Proline-Rich Region of Tau Is Regulated by Tau's N-Terminal Domain. *J. Biol. Chem.* **2019**. https://doi.org/10.1074/jbc.RA119.010172.
(183) Zhou, Z.; Roelfes, G. Synergistic Catalysis in an Artificial Enzyme by Simultaneous Action of Two Abiological Catalytic Sites. *Nat. Catal.* **2020**. https://doi.org/10.1038/s41929-019-0420-6.

(184) Burke, A. J.; Lovelock, S. L.; Frese, A.; Crawshaw, R.; Ortmayer, M.; Dunstan, M.; Levy, C.; Green, A. P. Design and Evolution of an Enzyme with a Non-Canonical Organocatalytic Mechanism. *Nature* **2019**. https://doi.org/10.1038/s41586-019-1262-8.
(185) Rogers, J. M.; Passioura, T.; Suga, H. Nonproteinogenic Deep Mutational Scanning of Linear and Cyclic Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2018**. https://doi.org/10.1073/pnas.1809901115.
(186) Rogers, J. M.; Oleinikovas, V.; Shammas, S. L.; Wong, C. T.; De Sancho, D.; Baker, C. M.; Clarke, J. Interplay between Partner and Ligand Facilitates the Folding and Binding of an Intrinsically Disordered Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014**. https://doi.org/10.1073/pnas.1409122111.
(187) Schönherr, H.; Cernak, T. Profound Methyl Effects in Drug Discovery and a Call for New C-H Methylation Reactions. *Angewandte Chemie - International Edition*. 2013. https://doi.org/10.1002/anie.201303207.
(188) Yang, B.; Wang, N.; Schnier, P. D.; Zheng, F.; Zhu, H.; Polizzi, N. F.; Ittuveetil, A.; Saikam, V.; Degrado, W. F.; Wang, Q.; et al. Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins. *J. Am. Chem. Soc.* **2019**. https://doi.org/10.1021/jacs.9b02611.
(189) Josephson, B.; Fehl, C.; Isenegger, P. G.; Nadal, S.; Wright, T. H.; Poh, A. W. J.; Bower, B. J.; Giltrap, A. M.; Chen, L.; Batchelor-McAuley, C.; et al. Light-Driven Post-Translational Installation of Reactive Protein Side Chains. *Nature* **2020**. https://doi.org/10.1038/s41586-020-2733-7.
(190) Doud, M. B.; Hensley, S. E.; Bloom, J. D. Complete Mapping of Viral Escape from Neutralizing Antibodies. *PLoS Pathog.* **2017**. https://doi.org/10.1371/journal.ppat.1006271.
(191) Li, Q.; Chen, Q.; Klauser, P. C.; Li, M.; Zheng, F.; Wang, N.; Li, X.; Zhang, Q.; Fu, X.; Wang, Q.; et al. Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics. *Cell* **2020**. https://doi.org/10.1016/j.cell.2020.05.028.
(192) Arasanz, H.; Gato-Cañas, M.; Zuazo, M.; Ibañez-Vea, M.; Breckpot, K.; Kochan, G.; Escors, D. PD1 Signal Transduction Pathways in T Cells. *Oncotarget*. 2017. https://doi.org/10.18632/oncotarget.17232.
(193) Arakhamia, T.; Lee, C. E.; Carlomagno, Y.; Duong, D. M.; Kundinger, S. R.; Wang, K.; Williams, D.; DeTure, M.; Dickson, D. W.; Cook, C. N.; et al. Posttranslational Modifications Mediate the Structural Diversity of Tauopathy Strains. *Cell* **2020**. https://doi.org/10.1016/j.cell.2020.01.027.